

सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI
56376

I.A.R.I.6.
GIP NLK—H-3 I.A.R.I. -10-5 55—15,000

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Volume XLVII, Art. 3. Pages 241–374

Editor
ROY WALDO MINER

Consulting Editor
JAMES B. ALLISON

Associate Editor
MICHAEL DEMAREST

PROTEINS AND PROTEIN HYDROLYSATES IN NUTRITION

BY

JAMES B ALLISON, JOHN A. ANDERSON, RICHARD H BARNES, DAVID K. BOSSHARDT, BACON F CHOW, CO TUI, ROBERT ELMAN, L. L MILLER, JOHN P PETERS, F S. ROBSCHEIT-ROBBINS ROBERT D SEELEY, AND G. H. WHIPPLE



576376

NEW YORK
PUBLISHED BY THE ACADEMY
OCTOBER 30, 1946

THE NEW YORK ACADEMY OF SCIENCES

(LYCEUM OF NATURAL HISTORY, 1817–1876)

OFFICERS, 1946

President: WALTER H. BUCHER

Vice-Presidents: CHARLES H. BEHRE, JR.; ROSS F. NIGRELLI; EMILY T. BURR; GEORGE HERZOG; JOSEPH S. FRUTON; ATHELSTAN F. SPILHAUS

Recording Secretary: CHARLES M. BREDER, JR.

Corresponding Secretary: H. HERBERT JOHNSON

Treasurer: HARDEN F. TAYLOR

Librarian: ERICH M. SCHLAIKJER

Editor: ROY WALDO MINER

Elected Councilors:

1944–1946: VICTOR K. LAMER; THEODORE SHEDLOVSKY; WILBUR G. MALCOLM

1945–1947: RALPH H. CHENEY; ROBERT CUSHMAN MURPHY; HERBERT F. SCHWARZ

1946–1948: HANS T. CLARKE; CHARLES GLEN KING; HANS MOLISCH

Finance Committee: W. REID BLAIR (*Chairman*); MARVIN D. THORN; ADDISON WEBB

OFFICERS OF SECTIONS

GEOLOGY AND MINERALOGY

Chairman: CHARLES H. BEHRE, JR. *Secretary:* ARTHUR N. STRAHLER

BIOLOGY

Chairman: ROSS F. NIGRELLI. *Secretary:* CHARLES R. SCHROEDER

PSYCHOLOGY

Chairman: EMILY T. BURR. *Secretary:* S. STANSFELD SARGENT

ANTHROPOLOGY

Chairman: GEORGE HERZOG. *Secretary:* WALTER DYK

PHYSICS AND CHEMISTRY

Chairman. JOSEPH S. FRUTON. *Secretary:* TURNER ALFREY, JR.

OCEANOGRAPHY AND METEOROLOGY

Chairman: ATHELSTAN F. SPILHAUS. *Secretary:* JOHN C. ARMSTRONG

The Sections of Geology and Mineralogy, Biology, Psychology, and Anthropology meet in regular session on Monday evenings at 8:00 o'clock from October to May, inclusive. The Sections of Physics and Chemistry, and Oceanography and Meteorology hold two-day conferences at irregular periods. All meetings of the Academy are held at The American Museum of Natural History, Central Park West at 79th Street, New York, N. Y.

Annals of the New York Academy of Sciences
Volume XLVII Art 3 Pages 241–371
October 30 1946

PROTEINS AND PROTEIN HYDROLYSATES IN NUTRITION

By

James B Allison John A Anderson, Richard H Barnes David K Bosshardt Bacon F Chow Co Tui Robert Elman I I Miller John P Peters T S Robscheit-Robbins Robert D Stetty and G H Whipple

CONTENTS

	PAGE
Introduction By James B Allison	243
The Determination of the Nitrogen Balance Index in Normal and Hypoproteinemic Dogs By James B Allison John A Anderson and Robert D Stetty	245
The Evaluation of Protein Quality in the Normal Animal By Richard H Barnes and David K Bosshardt	273
The Electrophoretic Studies on the Effect of Protein Depletion on Plasma Proteins and the Regeneration of Plasma Proteins after Oral Administration of Hydrolysates Prepared from Casein and Lactalbumin By Bacon F Chow	297
Blood Protein Regeneration and Interrelation By G H Whipple T S Robscheit Robbins and I I Miller	317
Nitrogen Metabolism in Acute and Chronic Disease By John P Peters	327
The Intravenous Use of Protein and Protein Hydrolysates By Robert Elman	345
Clinical Experiences with Oral Use of Protein Hydrolysates By Co Tui	359

This series of papers is the result of a Conference on Proteins and Protein Hydrolysates held by the Sections of Biology and Physics and Chemistry of the New York Academy of Sciences December 7 and 8 1945

Publication made possible through a grant from the Conference Publications Revolving Fund

COPYRIGHT 1946

BY

THE NEW YORK ACADEMY OF SCIENCES

ANNALS

OF THE

NEW YORK ACADEMY OF SCIENCES

VOLUME XLVII



NEW YORK
PUBLISHED BY THE ACADEMY
1946–1947

Editor
ROY WALDO MINER

Associate Editors
MICHAEL DEMAREST
LOTHAR SALIN

Consulting Editors
WILLIAM H. STEIN
(Pages 57–240)

JAMES B. ALLISON
(Pages 241–374)

DAVID NACHMANSOHN
(Pages 375–602)

S. BERNARD WORTIS
(Pages 603–664)

ALEXANDER SANDOW
(Pages 665–930)

CONTENTS OF VOLUME XLVII

PAGE

Title Page i

Contents iii

The Golgi Apparatus—An Interpretation of Its Structure and Significance. By Leonard G. Worley 1

Amino Acid Analysis of Proteins. By William H Stein, Reginald M. Archibald, Erwin Brand, R. Keith Cannan, Hans T. Clarke, John T. Edsall, G. L. Foster, Stanford Moori, David Shemin, Esmond E. Snell, and Hubert B Vickery 57

Proteins and Protein Hydrolysates in Nutrition. By James B. Allison, John A. Anderson, Richard H. Barnes, David K. Bosshardt, Bacon F. Chow, Co Tui, Robert Elman, L. L. Miller, John P. Peters, F. S. Robscheit-Robbins, Robert D. Seelly, and G. H. Whipple 241

The Physico-Chemical Mechanism of Nerve Activity. By David Nachmansohn, Charles M. Berry, Oscar Bodansky, Frank Brink, Jr, Detlev W. Bronk, M. Verlaine Brown, C. W. Coates, R. T Cox, J. C. Eccles, Alfred Fessard, J. F. Fulton, R. W. Gerard, Alfred Gilman, D. E. Green, Joseph C. Hinsey, Rudolph Höber, Martin G. Larrabee, and Tracy J. Putnam 375

Physiological and Psychological Factors in Sex Behavior. By S Bernard Worris, Gregory Bateson, William E Gait, Morris Herman, Alfred C Kinsey, and William C. Young 603

Muscular Contraction. By Alexander Sandow, R S Bear, Ernst Fischer, A S Gibson, Jr, Eugene Guth, C. F Hall, M. V. Jakus, S W. Kuffler, Otto Meyerhof, Severo Ochoa, Robert W Ramsey, F. O. Schmitt, G. M Scholpitt, Shih Chang Shen, H. Burr Steinbach, and S. M. Walker 665

INTRODUCTION

By James B. Allison

Bureau of Biological Research, Rutgers University

This conference on "Proteins and Protein Hydrolysates in Nutrition" is being held, I believe, on the threshold of a new era in nutrition. We have had the eras of calories, of vitamins, and now we find ourselves in the era of proteins and amino acids. In each of these periods, special attention has been given to these limited aspects of nutrition, but sufficient regard has not been given to the interrelation between them. We are now ready to give more attention to these interrelations, realizing that living systems must receive the proper proportions of all essential ingredients in the diet to maintain health, each ingredient having an important role to play, singly and in relation to others, in building, maintaining, and repairing the living machine. Thus, a new era of integration arises, correcting the distorted concepts of the past, and making possible the development of modern diet therapies, such as those to be discussed in this conference.

We shall center our attention on proteins and amino acids, at no time forgetting the relation of these important ingredients to others in the diet. We know that the amino acids work as members of teams of various combinations to build a hormone or an enzyme; to form a bit of tissue or an antibody to disease; to repair a peptic ulcer or to heal a wound. We shall study these amino acids, and the proteins which furnish them, to discover the job they do in maintaining life in normal tissues or in repairing damage to injured ones.

The war has taught us to cooperate in research and to look boldly into the future, building the new on the substantial contributions of the past. The biochemists, the physiologists, the clinicians, at this conference, look together at the field of protein and protein hydrolysates in nutrition, first at the over-all job the amino acids do in different combinations in building and in maintaining the living machine, and then at the application of this knowledge to modern protein therapy.

THE DETERMINATION OF THE NITROGEN BALANCE INDEX IN NORMAL AND HYPOPROTEINEMIC DOGS*

By James B. Allison, John A. Anderson, and Robert D. Seeley

Bureau of Biological Research, Rutgers University, New Brunswick, N. J.

Food protein nitrogen, entering into the dynamic equilibrium of the body, develops a characteristic nitrogen balance, reflecting the retention of nitrogen by the body. Recognizing the fundamental nature of the fraction of nitrogen retained in the body, Mitchell and others (1944) have called it the "biological value" of the protein. The significance of "biological value" (*BV*), correlated with nitrogen balance (*NB*) and absorbed nitrogen (*AN*), was illustrated by Allison and Anderson (1945), in the following equation:

$$NB = (BV)\ (AN) - EN, \qquad (1)$$

where *EN* is so-called body or endogenous nitrogen. Nitrogen balance (*NB*), absorbed nitrogen (*AN*), and endogenous nitrogen (*EN*) must be known, before the "biological value" (*BV*) can be calculated. Unfortunately, *EN* cannot be calculated with certainty. Schoenheimer and Rittenberg (1940), for example, have demonstrated a dynamic equilibrium between the tissues and the surrounding media which makes a separation of food nitrogen from body nitrogen a difficult, perhaps impossible, task. This complete intermingling of body nitrogen and of food nitrogen does not prevent, however, the expression of the utilization of food nitrogen in terms of the two fundamental variables, nitrogen intake and nitrogen output. The retention of nitrogen in the body is a function of these two variables, the relationship between which must be established to permit a logical approach to the subject of nitrogen metabolism. Given a knowledge of this relationship, nitrogen sources can be studied: (1) according to the nitrogen balances produced; and (2) according to specific nitrogen retention, as, for example, in the formation of plasma proteins. Studies such as these, using adult dogs as experimental animals, are in progress in our laboratories. The data obtained, both published and unpublished, on the correlations of nitrogen balance with nitrogen intake, in normal dogs, and with intake

* These studies were supported by the Protein Metabolism Fund of the Bureau of Biological Research, Rutgers University, and were undertaken, in part, in cooperation with the Quartermaster Corps Committee on Food Research.

and plasma protein regeneration, in hypoproteinemic dogs, are reviewed in the following pages.

METHODS

Adult dogs in good health, weighing from 6 to 10 kg. and made parasite-free, were used in these studies.

The protein-free diet, similar to one used by Melnick and Cowgill (1937), is recorded in TABLE 1. The daily caloric intake was adjusted to maintain an optimal nutritive condition as estimated by Cowgill (1928). This intake varied, according to the needs of each dog, from 60 to 100 calories per kilogram of body weight per day, and was kept constant during an experimental period. Proteins were included in the diet, by replacing equivalent amounts of calories from carbohydrate. When a natural foodstuff containing protein was added to the protein-free diet, adjustments were made for the carbohydrate and fat content of that foodstuff, the caloric intake being kept constant.

TABLE 1

Protein-Free Diet			*Vitamin Supplements*	
	Per Kilogram Body Weight			Mg. per Kilogram Body Weight Per Day
	Calories	Grams		
			Thiamin	0.025
			Riboflavin	.025
Sucrose	14.0	3.50	Nicotinic acid	.250
Dextrin	12.6	3.15	Calcium pantothenate	.200
Glucose	21.4	5.35	Pyridoxine	.015
Lard	32.0	3.55	Choline	15.0
Salt*	—	0.30	2-Methyl-Napthoquinone	.00001
Agar	—	0.40	Calcium α tocopherol monosuccinate	0.5
	—	—	Navitol (Squibb)	470 Units A
	80.0	16.25**		85 Units D

* Wesson's Modified Osborne-Mendel Salt Mixture.
** Mix 1.4 grams of water with every gram of dog food.

In the early experiments, the rations were fed for eight days, feces (with carmine markers) and urine being collected for analysis the last four days. In later experiments, the adjustment period was sometimes reduced from four to three or even two days, and the collection period to three days. The feces and urine samples, kept separate as daily samples or pooled for the collection period, were analyzed for nitrogen by the micro-Kjeldahl method, using selenium oxychloride as the catalyst.

The dogs, for plasma protein regeneration studies, were depleted by feeding the protein-free diet at a constant and adequate caloric intake. Two days of fasting, followed by two or three days of plasmapheresis, were usually used to speed up the rate of depletion. After this, the dogs received the protein-free diet, until the plasma protein concentration was reduced to approximately 4.0 gms. per cent and was essentially constant (Seeley, 1945).

Plasma volumes were determined, using the dye T-1824, according to the technique described by Gregersen and Stewart (1939). Plasma proteins were separated into "albumin" and "globulin" fractions, by the method of Howe (1921), as modified by Robinson, Price, and Hogden (1937).

RESULTS

The relationship between nitrogen balance (NB) and absorbed protein nitrogen (AN) in adult dogs was shown by Allison and Anderson (1945) to be linear in the region of negative and low positive nitrogen balance. The equation for this relationship is:

$$NB = K\ (AN) - NE_0, \qquad (2)$$

where K is the slope of the line and NE_0 the excretion of nitrogen on the protein-free diet.* EQUATION 2 describes, for example, the linear portion of the curve drawn through the white circles in FIGURE 1A. These circles, correlating absorbed nitrogen with nitrogen balance, are averaged data, obtained while feeding casein to three dogs. The y intercept or NE_0 is 2.54 gms/day/sq.M. of body surface area; absorbed nitrogen at equilibrium is 3.1 gms/day/sq.M.; and K, the slope of the straight line, is 0.82. This slope, the rate of change of nitrogen balance with respect to absorbed nitrogen, is a function of the retention of nitrogen, and is the "biological value" (BV), if NE_0 in EQUATION 2 equals EN in EQUATION 1. The value of K, however, need not be identical with BV. It is suggested, therefore, that K be called the nitrogen balance index, rather than the "biological value" of the protein. The line in FIGURE 1A becomes obviously curvilinear with increasing nitrogen intake on the positive side of nitrogen balance. Thus, the nitrogen balance index (K), which is constant in the region of negative or low positive nitrogen balance, is a decreasing variable at higher nitrogen intakes.

* After this review was submitted for publication, a paper by **Bricker, M., H. H. Mitchell, & G. M. Kinsman** (1945) proved that this same relationship described data obtained in man.

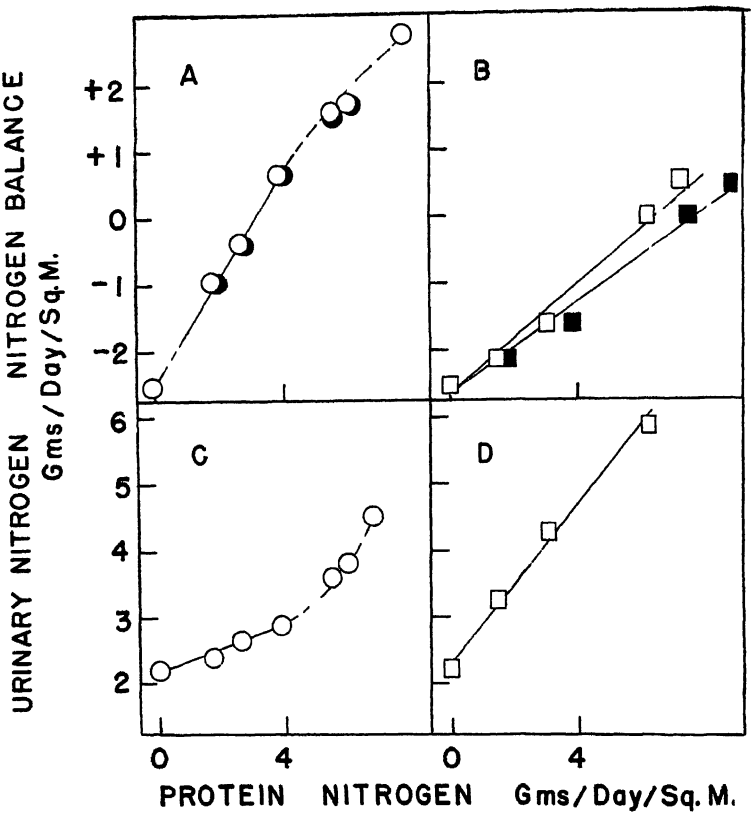


FIGURE 1 Nitrogen intake in gms/day/sq M, plotted against nitrogen balance (A, B) or urine nitrogen (C, D), in gms/day/sq M of body surface These are average data obtained on three dogs (A) The white circles correlate absorbed casein nitrogen the black circles, total casein nitrogen intake with nitrogen balance (B) The white squares correlate absorbed nitrogen, the black squares total nitrogen intake, with nitrogen balance for experiments on the three dogs receiving a derived soybean protein (protein A) (C) The white circles correlate absorbed casein nitrogen with urine nitrogen, in the same experiment described in (A) (D) The white squares average data obtained on the protein derived from soybean (data taken from **Allison and Anderson,** 1945)

Absorbed nitrogen (AN), in EQUATION 2, is calculated from the total nitrogen intake (NI) as follows:

$$AN = D \times NI, \tag{3}$$

where D is true digestibility. Since D is essentially constant for any one protein, then

$$NB = K' (NI) - NE_0, \tag{4}$$

where K' is the rate of change of nitrogen balance, with respect to total nitrogen intake (NI). The black circles, in FIGURE 1A, correlate NI with NB, so that a line drawn through these circles would have a slope

equal to K'. Since the digestibility of casein is 0.96, there is little difference between K and K' for casein

There is a more marked difference between K and K', in FIGURE 1B, which illustrates data obtained while feeding a protein derived from soybean (protein A). In this figure, the white squares correlate absorbed nitrogen, and the black squares, total nitrogen intake, with nitrogen balance. The slope of the line drawn through the white squares is 0.4 (K) and through the black squares is 0.33 (K'). Since K and K' are correlated through digestibility (D) as follows,

$$D = \frac{K'}{K}. \tag{5}$$

K' is a function of digestibility, as well as of the retention of protein nitrogen in the body of the animal.

Another fundamental expression of the relationship between nitrogen input and output was derived from EQUATION 2, by Allison and Anderson (1945), as being

$$UN = (1\text{-}K)\ (AN) + UN_0, \tag{6}$$

where UN is urine nitrogen excretion, UN_0 is urine nitrogen excretion on a protein-free diet, and the other symbols are the same as those used previously. Where EQUATION 2 describes a linear relationship between nitrogen balance and absorbed nitrogen, EQUATION 6 describes a linear relationship between urine nitrogen and absorbed nitrogen. In FIGURE 1C, urine nitrogen in gms/day/sq.M. is plotted against absorbed nitrogen in gms/day/sq M. of body surface, averaging data from the same experiments on casein as are illustrated in FIGURE 1A. Similarly, the squares in FIGURE 1D average urine nitrogen data from the same experiments as are described in FIGURE 1B. The marked curvilinear nature of the relationship between urine nitrogen and absorbed nitrogen, when the animal is in positive nitrogen balance, is illustrated by the data on casein. In this region of positive balance, the rate of change of urine nitrogen output, with respect to nitrogen input, is increasing rapidly, while K, the nitrogen balance index, is decreasing Since all the points plotted in FIGURE 1D are in, or near, the region of negative nitrogen balance, the curvilinear portion was not reached in these data

FIGURE 2 summarizes data on the effects of caloric intake on the utilization of protein nitrogen by adult dogs. Dried, uncooked egg al-

* The authors thank Dr Samuel Lepkovsky, of the Military Planning Division of the Office of the Quartermaster General, for suggestions and advice in these studies on caloric intake.

bumin* was the source of protein nitrogen, in these experiments. The circles represent average data obtained on two dogs receiving an adequate caloric intake; the rectangles average data obtained on three dogs receiving fifty per cent of their normal caloric intake; and the triangles, average data obtained on three dogs receiving twenty-five per cent of normal intake. The nitrogen balance index (K) is 0.96 for normal adult dogs receiving an adequate or half-adequate caloric intake. The

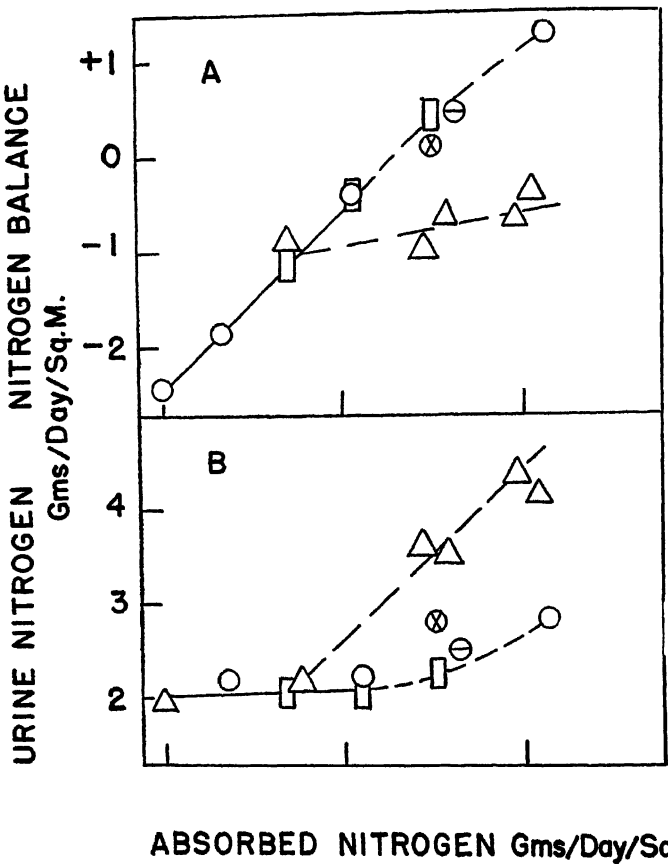


FIGURE 2. Absorbed egg albumin nitrogen in gms/day/sq.M. are plotted against nitrogen balance (A) or urine nitrogen (B) in gms/day/sq.M. of body surface. The circles represent average data obtained on two dogs receiving an adequate caloric intake; the rectangles, average data on three dogs receiving fifty per cent of their normal caloric intake; and the triangles, average data on three dogs receiving twenty-five per cent of an adequate caloric intake. The circle with the bar illustrates the effect of adding sucrose to increase the caloric intake from twenty-five to fifty per cent of normal. The circle with the cross represents similar data, obtained by increasing the caloric intake with lard. (B) The same experiments which are averaged in (A) are illustrated, also, through the correlation between absorbed nitrogen and urine nitrogen excretion.

* Dried, uncooked egg albumin, manufactured by Swift & Company.

caloric intake did not affect the utilization of egg white, as measured by K, until the calories were reduced to twenty-five per cent of normal. With this low caloric intake, the dogs were not put into positive nitrogen balance, even with nitrogen intakes as high as 4 gms/day/sq.m. of body surface. Indeed, there was very little nitrogen, utilized by the dogs, above 1.5 gms/day/sq.M. absorbed nitrogen, the excess appearing in the urine. These data demonstrate that utilization of protein is controlled by the presence in the diet of a critical proportion of calories.

The circle with the horizontal bar, in FIGURE 2, averages data obtained when the caloric intake was increased again to half normal requirements with sucrose, and the circle with the cross represents data obtained by increasing the caloric intake an equivalent amount with lard. The protein-sparing actions of carbohydrate and of fat are nicely demonstrated thus.

TABLE 2 summarizes some of the data obtained while feeding different protein sources to normal adult dogs. Under absorbed nitrogen at equilibrium, in the second column, are values which represent the protein minima for nitrogen equilibrium. Two groups of values for casein are recorded to emphasize that protein minimum may vary, while K or K' remains constant. The protein minimum of 4.0 gms/day/sq.M., for

TABLE 2

PROTEIN SOURCES, ABSORBED NITROGEN AT EQUILIBRIUM, K (EQUATION 2), K' (EQUATION 4), AND "NET PROTEIN" (K' × FRACTION OF PROTEIN IN SOURCE) DATA OBTAINED FEEDING THE PROTEIN SOURCE TO NORMAL ADULT DOGS.

Each set of data represents averages from experiments on three dogs.

Protein Source	Absorbed Nitrogen at Equilibrium	Nitrogen Balance Indices		"Net Protein"
		K	K'	
	gms/day/sq. M.			
Lactalbumin	2.2	1.0	0.97	0.74
Dried egg white	2.5	0.96	0.91	0.63
Casein	4.0	0.82	0.78	0.60
Casein	3.1	0.81	0.77	0.59
Squibb casein hydrolysate*	2.2	0.80	0.79	0.60
Amigen†	2.2	0.80	0.79	0.60
Chicken entrails	3.3	0.77	0.73	0.09
Flounder entrails	4.3	0.77	0.74	0.11
Flounder heads	6.4	0.52	0.43	0.06
Protein A‡	6.6	0.39	0.32	0.28

* This is a casein hydrolysate prepared in the Squibb Institute for Medical Research by digestion of casein with trypsin.
† Amigen (batch 10067) was furnished by Mead Johnson Company.
‡ Derived from soybean.

the first group of three dogs receiving casein, is greater than the minimum of 3.1 for the second group, a difference which can be interpreted to mean that the first group was excreting more nitrogen from body stores than was the second. The first group, therefore, required more absorbed protein nitrogen to maintain equilibrium. Thus, protein minima for any one protein may vary, while K or K' is not altered. These nitrogen balance indices are not easily affected by changes in the physiological state of the animal and are more fundamental functions of retention of protein nitrogen by the animal than are protein minima. These observations agree with those of Melnick and Cowgill (1937), who reported that the relationship between nitrogen balance and per cent protein caloric intake tended to be represented by parallel lines, while protein minima varied. The "net protein" values recorded in the last column of the table were calculated by multiplying K' by the fraction of protein in the protein source. Net protein values, the use of which was suggested by Mitchell and Carman (1924), are functions of the retention of nitrogen by the animal, digestibility, and concentration of protein in the source. Net protein values measure, therefore, the over-all efficiency of a protein source as a supply of nitrogen to the animal.

The same relationships between nitrogen balance and nitrogen intake (EQUATIONS 2 and 3) were found in hypoproteinemic, as in normal, dogs (Allison, Seeley, Brown, and Anderson, 1946). The values, however, for K and K' are higher in the hypoproteinemic, than in the normal, dog; these differences are illustrated in the last two columns of TABLE 3. The data in this table demonstrate that, as the plasma proteins are reduced in concentration, absorbed nitrogen at equilibrium and NE_0 tend to decrease, while nitrogen balance indices increase. The altered physiological state of the hypoproteinemic dog increases, therefore, the utilization of protein nitrogen.

The data in FIGURE 3 illustrate correlations between values for K and the daily excretion of urine nitrogen in normal and depleted dogs. The white circles record data obtained while feeding the protein-free diet, the black circles, while feeding protein nitrogen. Curve I illustrates data obtained on a normal dog which received Squibb casein hydrolysate nitrogen equivalent to 1.84 gms/day/sq.M. of body surface for eight days. Since the nitrogen balance index (K) for this hydrolysate is less than unity (0.8), the excretion of urine nitrogen increased during the nitrogen feeding period. Curve II was obtained while feeding

TABLE 3

THE EFFECT OF DEPLETION IN PROTEIN, REFLECTED IN A DECREASE FROM NORMAL IN PLASMA PROTEIN CONCENTRATION, UPON VALUES FOR ABSORBED NITROGEN AT EQUILIBRIUM, NITROGEN EXCRETION ON A PROTEIN-FREE DIET, K AND K' IN EQUATIONS 2 AND 4, RESPECTIVELY

Casein, Squibb casein hydrolysate,‡ Amigen,† Aminoids, and a protein derived from soybean (Protein A) were used as sources of nitrogen (**Allison, Seeley, Brown, & Anderson, 1946**).

Dog No.	Plasma Protein	Absorbed Nitrogen at Equilibrium	Nitrogen Excretion on Protein-Free Diet NE_0	Nitrogen Balance Indices	
				K	K'
	gms/100 ml.	gms/day/sq.M.	gms/day/sq M.		
		Casein			
*	6.2	3.20	2.59	0.81	0.77
31	5.6	2.55	2.00	0.79	0.70
44	4.6	1.65	1.47	0.89	0.83
42	4.6	1.55	1.39	0.90	0.88
55	4.1	2.06	1.90	0.92	0.92
28	1.1	1.73	1.61	0 93	0.93
		Squibb Casein Hydrolysate			
44	6.0	2.15	1.72	0.80	0.76
48	5.7	2.94	2.32	0.79	0.79
31	5.0	1.49	1 39	0.93	0.90
		Amigen			
44	6.0	2.15	1.72	0.80	0.76
48	5.2	2.72	2.28	0.84	0.84
31	4.8	1.90	1.77	0.93	0.93
		*Aminoids***			
60	6.0	4.60	2.20	0.48	0.40
48	5.0	3.50	2.10	0.60	0.52
55	3.9	2.42	1.69	0.70	0.60
		Protein A			
*	6.2	6.60	2.50	0.39	0.32
50	1.1	2.90	1.65	0.57	0.51
42	1.0	1 50	1.10	0 73	0.66

‡ See footnotes, TABLE 2.
* Data previously published (**Allison & Anderson, 1945**).
** Aminoids (batch 2851) was furnished by the Arlington Chemical Company.

egg albumin nitrogen equivalent to 1.44 gms/day/sq.M. for four days. The nitrogen balance index (K) for egg albumin was unity, in this experiment; hence, the excretion of urine nitrogen was independent of the nitrogen fed. Curve III illustrates data obtained on a hypoproteinemic dog which received egg albumin nitrogen equivalent to 1.52 gms/day/sq.M. of body surface. The nitrogen balance index (K), in this ex-

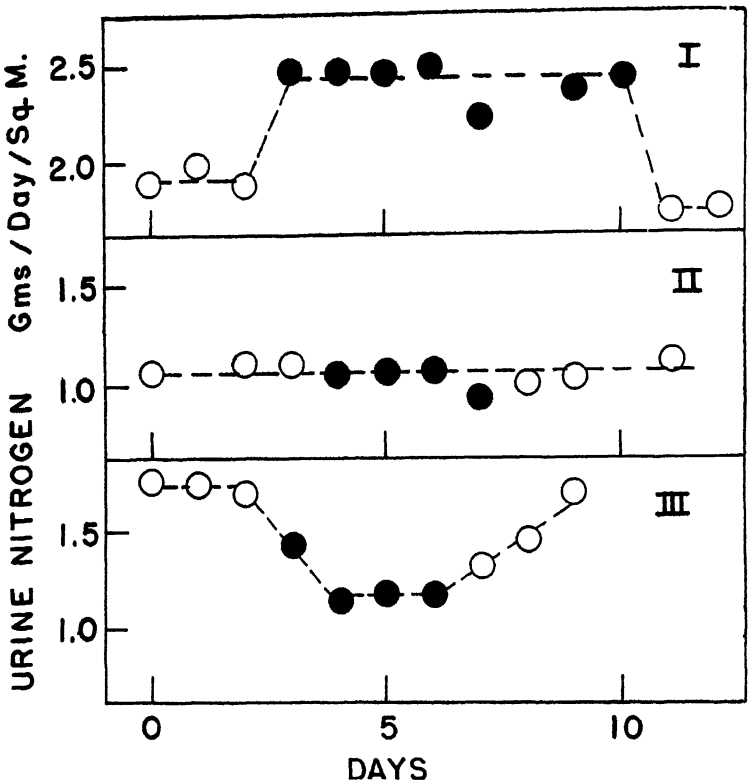


FIGURE 3. Urine nitrogen in gms/day/sq.M. of body surface, plotted against time in days. The white circles record data obtained while feeding the protein-free diet to a dog; the black circles, while feeding protein nitrogen. The black circles in I represent data obtained on a normal dog, while feeding 1.84 gms. of Squibb casein hydrolysate nitrogen/day/sq.M.; the black circles in II, while feeding 1.44 gms. of egg albumin nitrogen/day/sq.M.; the black circles in III represent data obtained on a hypoproteinemic dog, fed 1.52 gms. of egg albumin nitrogen/day/sq.M. (**Allison, Seeley, Brown, & Anderson,** J. Nutrition, 1946.)

periment, was greater than unity, the urine nitrogen excreted during the nitrogen feeding period being less than during the protein-free feeding period.

These data on the daily excretion of nitrogen on hypoproteinemic dogs fed egg albumin nitrogen illustrate again that the nitrogen balance index (K) increases in the depleted dog. More complete data on the utilization of egg albumin nitrogen by the normal and hypoproteinemic dog are presented in TABLE 4. The data obtained, using dogs 54, 50, and 31, are arranged chronologically, in this table, each determination of NB and NE_0 requiring about eight days. Values for K greater than unity were obtained for egg albumin fed to these hypoproteinemic dogs. Continued feeding of egg albumin resulted in a reduction in the value

TABLE 4

THE EFFECT OF PROTEIN DEPLETION, REFLECTED IN A DECREASE FROM NORMAL IN PLASMA PROTEIN CONCENTRATION, UPON NITROGEN BALANCE (*NB*), NITROGEN EXCRETION ON PROTEIN-FREE DIET (NE_0), *K* IN EQUATION 2

(Data taken from paper by Allison, Seeley, Brown, & Anderson, 1946.) Dried egg albumin was used as the source of nitrogen.

Dog No.	Plasma Protein	Absorbed Nitrogen *AN*	Nitrogen Balance *NB*	Nitrogen Excretion on a Protein-Free Diet NE_0	Nitrogen Balance Index *K*
	gms/100 ml.	gms/day/sq.M.	gms/day/sq.M.	gms/day/sq.M.	
*	6.0	2.50	0	2.55	0.96
54	5.4	1.87	+0.19	1.81	1.07
	5.0	0.89	−0.48	1.63	1.29
	5.0	1.44	+0.01	1.47	1.03
50	4.7	1.12	−0.56	1.91	1.20
	4.9	0.84	−0.58	1.49	1.08
	4.9	1.44	−0.04	1.45	0.99
31	4.9	1.63	+0.30	1.77	1.27
	4.7	0.85	−0.37	1.79	1.67
	4.7	1.26	−0.17	1.46	1.02

* Average data obtained on two normal dogs.

of NE_0 from approximately 1.8 to 1.42 gms/day/sq.M., values for *K* also being reduced to approximately unity. These data can be interpreted to mean that egg albumin spares body nitrogen. A similar conservation of body nitrogen by egg albumin has been reported by Willman, Swanson, Stewart, Stevenson, and Brush (1945) in some experiments on rats.

The sparing of body nitrogen can be illustrated by solving EQUATIONS 1 and 2 simultaneously for any given value of *NB* to yield:

$$EN = NE_0 - AN\ (K - BV), \qquad (7)$$

where the symbols have the same meaning as in the previous equations. According to EQUATION 7, if the nitrogen balance index *K* is equal to *BV*, then *EN* has the same value as NE_0, but, if *K* is greater than *BV*, then the excretion of endogenous nitrogen (*EN*) is less than the excretion of nitrogen on a protein-free diet (NE_0). A value for *BV* greater than unity is impossible. Since *K* is greater than unity in some of the egg albumin data, the feeding of this protein spares body nitrogen, the excretion of *EN* becoming less than NE_0.

Body nitrogen-sparing action is also illustrated by some data obtained while feeding methionine to three normal dogs receiving an adequate caloric intake. These data, averaged in FIGURE 4, demonstrate that the average urine nitrogen excretion, during an initial four day protein-

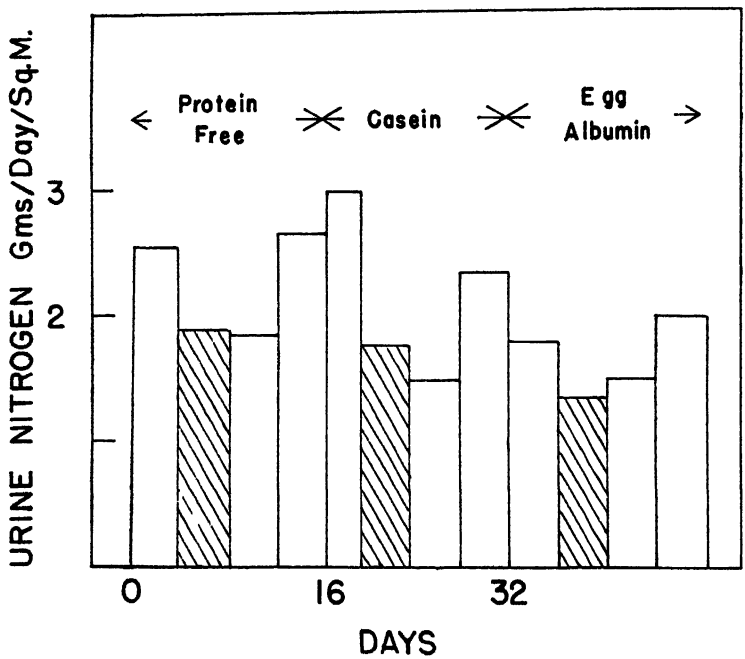


FIGURE 4 Urine nitrogen in gms/day/sq M. are plotted against time in days. These are averaged data obtained while feeding three dogs the protein-free diet followed by 1 8 gms of casein nitrogen/day/sq M then by 2 2 gms of egg albumin nitrogen/day/sq M One gram of *dl*-methionine was added daily to the diet of each dog, during the four day periods represented by the blocks containing slanted lines These data will appear in more detail in a paper by **Allison, Anderson, & Seeley** (in press)

free feeding period, was 2 6 gms/day/sq M. of body surface area. One gram of *dl*-methionine was fed daily to each dog for the next four days. The average excretion of urine nitrogen during this methionine-feeding period (lined block in FIGURE 4) was 1.9 gms/day/sq.M. Following these four days of amino acid feeding, the methionine was removed from the diet, the dogs receiving only the protein-free ration again The excretion of urine nitrogen remained lowered (1.8 gms/day/sq.M.), for four days after the removal of methionine, while the average excretion of urine nitrogen returned to the initial control level (2.7 gms/day/sq.M.), during the second four day period of protein-free feeding Thus, methionine spared body nitrogen, the sparing action lasting sev-

eral days after the amino acid had been removed from the diet. These results are in agreement with those recorded by Miller (1944), who found a decrease in the excretion of urine nitrogen in dogs, following the addition of methionine to a protein-free diet.

The experiment was repeated, after adding to the diet 1 8 gms. of casein nitrogen/day/sq.M. Since the nitrogen balance index (K) of casein is less than unity (0.8), the excretion of urine nitrogen, during the casein feeding period, was increased to 3 0 gms/day/sq M., which was greater than the excretion on the protein-free diet. The addition of 1 gm of *dl*-methionine, daily for four days, to the casein diet lowered the excretion of urine nitrogen to 1.8 gms/day/sq.M., sparing the body nitrogen of the dog. This sparing action of methionine was carried over again into the subsequent period, when the amino acid was not added to the casein diet.

The casein was replaced then with 2.2 gms. of egg albumin nitrogen/day/sq M. of body surface. The urine nitrogen decreased below that excreted while feeding casein, a result which is predicted from the higher nitrogen balance index for albumin than for casein. The addition of 1 gm of *dl*-methionine daily for four days to the egg albumin diet lowered the excretion of urine nitrogen to 1.35 gms/day/sq.M., revealing a sparing action, as in the casein experiment. Similarly, this sparing action was carried over into subsequent periods, when the methionine had been removed. Further experiments are being conducted to determine the supplementing value of methionine in casein or in egg albumin, and to determine the minimum amount of methionine needed to produce the sparing action demonstrated by the experiments in FIGURE 4.

Possibly, a different type of sparing action is illustrated by the reduction in the excretion of urine nitrogen which accompanies hypoproteinemia. This reduction is demonstrated by the data in TABLE 5, where the excretion on a protein-free diet of total nitrogen (UN_0), ammonia plus urea, and α-amino nitrogen are recorded before and after depletion. The average plasma protein concentrations in the control and depleted states were 6.0 and 4.2 gms. per 100 ml., respectively. The total urine nitrogen excretion (UN_0) was reduced from an average control of 2 95 gms/day/sq.M. to 1.92 gms/day/sq M. in the depleted state. Eighty-nine per cent of the decrease was in the excretion of the urea plus ammonia fraction, the α-amino nitrogen excretion not being altered. These data can be interpreted to mean that depletion in tissue proteins is reflected by a decrease in the use

TABLE 5

TOTAL AMMONIA AND UREA, AND α-AMINO URINE NITROGEN EXCRETION ON A PROTEIN-FREE DIET BEFORE AND AFTER DEPLETION

Dog No.	Control			Protein-Depleted		
	Total Nitrogen	Ammonia and Urea Nitrogen	α-Amino Nitrogen	Total Nitrogen	Ammonia and Urea Nitrogen	α-Amino Nitrogen
	gms/day/ sq.M.	gms/day/ sq.M.	gms/day/ sq.M.	gms/day/ sq.M.	gms/day/ sq.M.	gms/day/ sq.M.
28–1	3.75	2.91	0.11	1.50	1.09	0.11
21–6	3.00	2.36	0.08	2.58	1.94	0.13
30–1	3.32	2.48	0.22	1.66	1.08	0.14
28–2	2.00	1.37	0.04	1.76	1.29	0.04
27–1	3.13	2.66	—	1.75	1.07	—
21–4	2.49	1.67	—	2.24	1.42	—
Ave.	2.95	2.24	0.11	1.92	1.32	0.11

of protein stores, as measured through the excretion of waste products like urea and ammonia.

Depletion does not produce, however, a uniform reduction in all protein stores. Through electrophoretic analysis, Chow, Allison, Cole, and Seeley (1945) have demonstrated that depletion by protein-free feeding (hastened initially by several days of plasmapheresis) markedly reduces the total circulating plasma albumin in dogs, and, slightly, the total circulating β- and γ-globulin. However, it does not decrease, but even increases, the total circulating α-globulin. Similarly, Weech (1938/39), and Madden and Whipple (1940) have proved that the plasma albumin fraction is more markedly reduced than the globulin fraction, when dogs are depleted in proteins. The marked reduction in plasma volume which can accompany the reduction in total circulating albumin is illustrated in FIGURE 5, where circulating albumin in grams and plasma volume in milliliters are plotted against days of depletion. The dog is, therefore, depleted differentially in these circulating tissue proteins, a differential which should be found in other tissues, thus markedly altering the physiological state of the animal.

Using this technique of depletion, dogs were reduced to a constant state of hypoproteinemia, as illustrated by the first three points in FIGURE 6. The total plasma proteins in this dog were reduced to 4.0 gms., the albumin to less than 2.0 gms., while the total globulins were

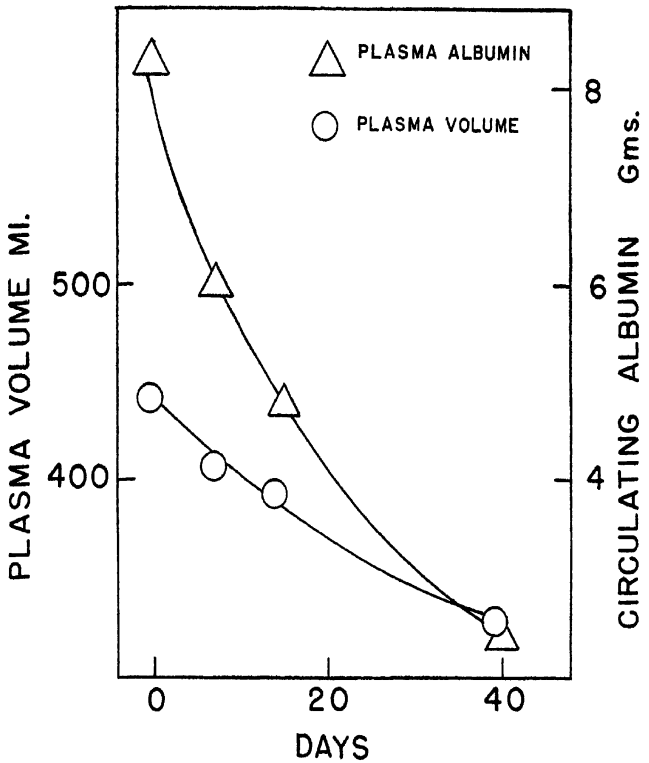


FIGURE 5. Plasma volume in ml. and circulating albumin in gms. are plotted against days of depletion.

but little reduced. The albumin and globulin concentrations were determined by salt fractionation (see Methods). At zero days, this dog was fed 6.6 gms. of lactalbumin hydrolysate[4] nitrogen/day/sq.M. of body surface. The rapid rate of increase of the total protein, during the first five days of feeding, followed by a slower rate of increase, is illustrated in the upper curve. The albumin fraction increased similarly to the total proteins, while the globulins showed an initial rise, followed by a decrease in concentration. Since the plasma volume, also, began to increase, sometime after the fifth day of protein hydrolysate feeding, much of the decrease in globulins can be attributed to dilution. Because of this dilution, changes in concentration of

[4] The lactalbumin hydrolysate, prepared by enzymatic hydrolysis, was furnished by Dr. Bacon F. Chow, of the Squibb Institute for Medical Research. Details of preparation will be published elsewhere.

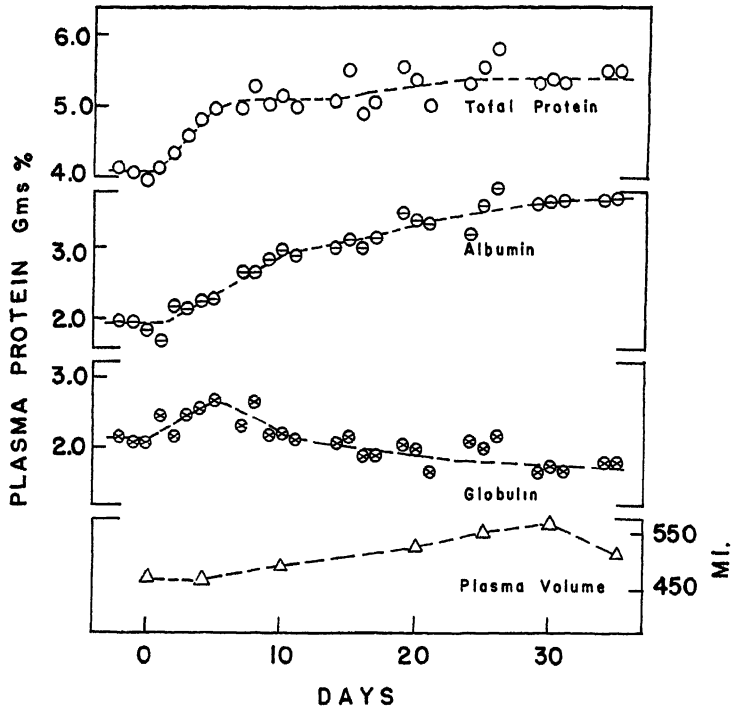


FIGURE 6. Data obtained on an hypoproteinemic dog receiving 6.6 gms. of lactalbumin hydrolysate nitrogen/day/sq.M. of body surface area, for thirty days. The white circles in the uppermost curve illustrate the effect of feeding the hydrolysate on the total proteins, below which are recorded changes in albumin (circles with bars) and in globulins (circles with crosses). The triangles record changes in plasma volumes.

plasma proteins beyond the fifth day of protein feeding do not reflect accurately the regeneration of total circulating proteins.

The data plotted in FIGURE 7 demonstrate that albumin/globulin ratios, determined by salt fractionation (Howe, 1921), are always greater, in a regular way, than those determined through electrophoresis.* The electrophoretic method has the advantage of yielding a more complete separation into what may be considered as absolute fractions (Chow, Allison, Cole, and Seeley, 1945), but, since the correlation between the ratios determined by the two methods is regular and essentially linear, either method reveals relatively similar changes. For reasons of economy and speed, the salt fractionation method was used to determine changes in albumin and globulin fractions in the following studies on regeneration of plasma proteins in hypoproteinemic dogs.

* The electrophoretic analyses were done by Dr Bacon F. Chow, of the Squibb Institute for Medical Research, during cooperative experiments which will be described elsewhere.

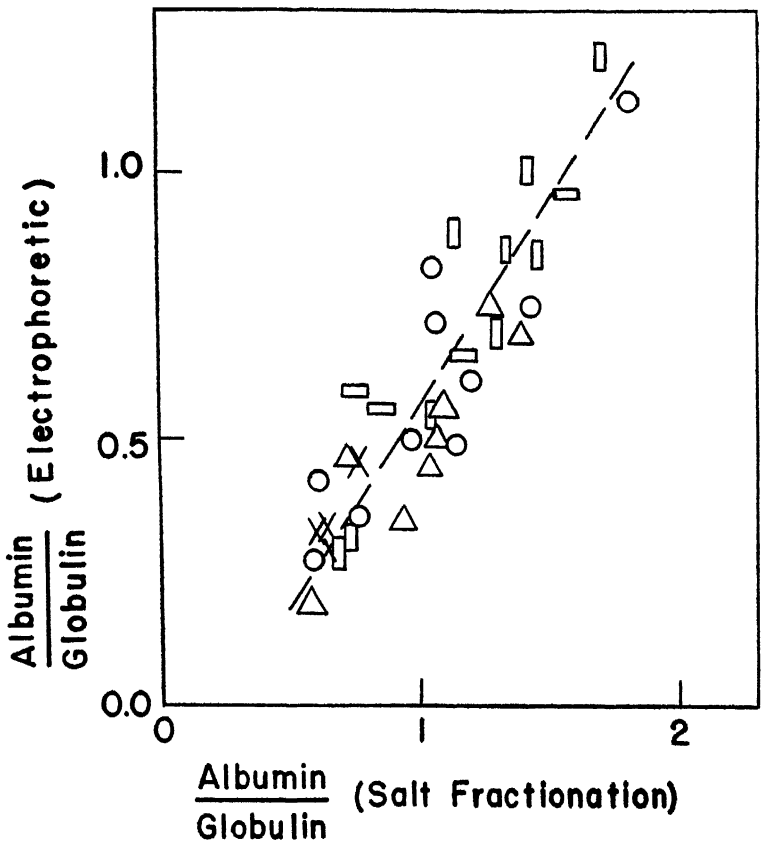


FIGURE 7. Albumin/globulin ratios determined by the electrophoretic method are plotted against those determined in the same plasma by salt fractionation. (**Howe, 1921.**)

Since the most rapid regeneration usually took place during the first five days of protein feeding, without changes in plasma volumes, the following technique was developed for these regeneration studies (Seeley, 1945). The dogs were depleted by feeding the protein-free diet and by a few days of plasmapheresis, until the plasma protein concentration was between 4.0 and 4.5 gms. per 100 ml. plasma and essentially constant. After the plasma protein had been reduced to this constant state of depletion, the dogs were fed protein in the diet for a period of five days. If the dogs were in positive nitrogen balance, during this feeding period, regeneration of plasma protein took place. After five days of protein feeding, the animals were returned to the protein-free diet. Thus, the dog was taken from a depleted state, through a regeneration period, back to the depleted state.

FIGURE 8 illustrates a regeneration experiment of the type just described, in which bovine serum protein was the source of dietary nitrogen. The dog received orally 6.6 gms. of protein nitrogen/day/sq.M. of body surface, which put the animal into a positive nitrogen balance of +2.5 gms/day/sq.M The average plasma protein concentration, be-

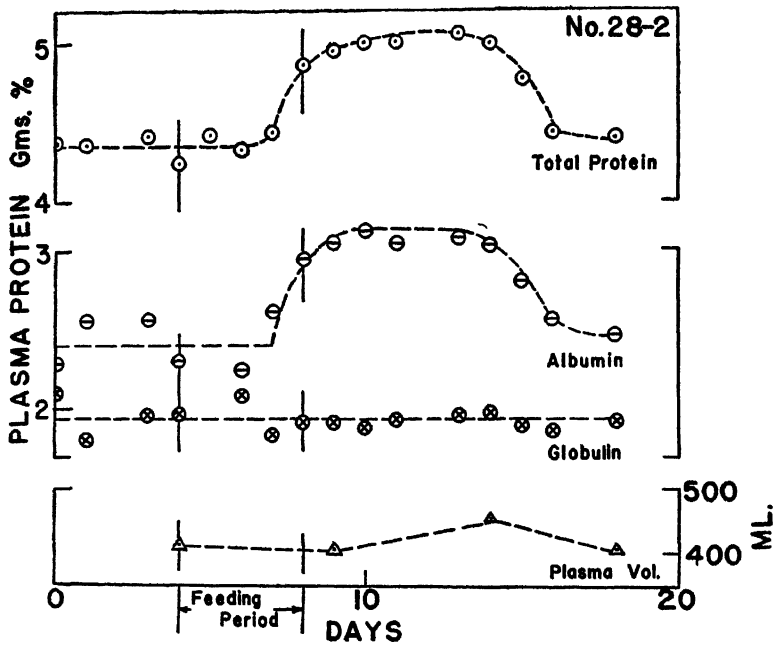


FIGURE 8 The effect of the oral feeding of 6 6 gms of bovine serum protein nitrogen/day/sq M. upon the plasma protein regeneration, in a hypoproteinemic dog The plasma protein concentration in gms per cent and the plasma volume in ml are plotted against time in days. The dog was fed the protein-free diet, from days 0 to 3, bovine serum protein, from days 4 to 8 ("Feeding Period"), and the protein-free diet, from day 9 to the end of the experiment (**Seeley, 1915**).

fore nitrogen feeding, was 4.34 gms. per cent. After the fourth day of protein feeding, the plasma protein concentration showed a definite rise, reaching a maximum of 5.10 gms. per cent two days later. The concentration remained elevated for five days, and then dropped quickly to the initial level of depletion.

The concentration of plasma albumin and globulins, determined by the salt fractionation method of Howe (1921), are illustrated below the curve for total protein in FIGURE 8. According to these data, the increase in plasma protein, as a result of feeding bovine serum protein, was due to a rise in albumin, but not in the globulin fraction. Since there were no marked changes in plasma volume, during this experi-

ment, the curves drawn through the points of concentration give a true picture of shifts in total circulating proteins.

The lag period, in protein regeneration following feeding of the bovine serum protein, was noted often in experiments of this type. Holman, Mahoney, and Whipple (1934) and Melnick, Cowgill, and Burach (1936) noted similar lag periods, after feeding protein to depleted dogs. Weech, Goettsch, and Reeves (1935) estimated that, during depletion by protein-free feeding alone, the loss in plasma protein represented only four per cent of the total loss of nitrogen by the dog. Elman and Davey (1943) injected whole dog plasma intravenously into depleted dogs and found that only ten to fourteen per cent remained in the blood stream, after twenty-four hours. Weech and Goettsch (1938) observed that no lag period occurred, after the initial protein intake, in mildly depleted dogs. These results confirm the conclusion that the larger protein reserves of the body must be partially replenished, before the plasma protein concentration will be increased.

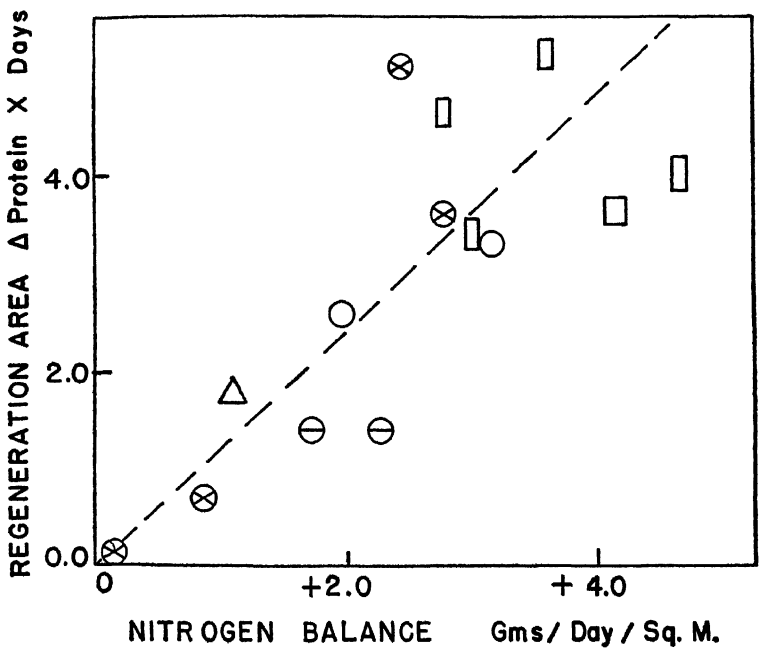


FIGURE 9. Regeneration areas (Δ protein × days) are plotted against positive nitrogen balances in gms/day/sq M. of body surface. The data were obtained on depleted dogs, orally fed bovine serum protein (circles with crosses), casein (circles), Squibb casein hydrolysate (rectangles), protein hydrolysate B (circles with bars), and Amigen (square). The triangle represents data obtained on a depleted dog fed Amigen intravenously (**Allison, Seeley, Brown, & Anderson,** J. Nutrition, 1946).

Regeneration of the plasma proteins was measured by calculating the areas under the curves showing an increase above the depleted concentration, as in FIGURE 8. These areas ($\triangle$ protein $\times$ days) take into consideration, not only the magnitude of the rise in plasma protein, but also the rate of depletion following nitrogen feeding, giving better correlation with nitrogen balances than could be obtained with the increase in plasma protein concentration alone.

The data plotted in FIGURE 9 prove that regeneration of plasma proteins, measured by these areas, increases in magnitude as positive nitrogen balance increases. Thus, regeneration of plasma proteins, in hypoproteinemic dogs, is a function of the nitrogen balance produced and, therefore, of the nitrogen balance index (K).

The kinds of plasma proteins regenerated, however, depend in part upon the pattern of amino acids fed. TABLE 6 records data illustrating

TABLE 6

THE PLASMA ALBUMIN AND GLOBULIN CONCENTRATIONS IN PROTEIN-DEPLETED DOGS, THE SOURCE OF PROTEIN FED FOR A PERIOD OF FIVE DAYS TO THE DEPLETED DOGS, THE NITROGEN BALANCES OBTAINED, AND REGENERATION AREAS

The regeneration area was taken to be the area under curves (gms % of plasma protein *vs.* days) rising above the plasma protein concentration in the depleted state (Seeley, 1945).

Depleted State		Protein Source	Nitrogen Balance	Regeneration Area ($\triangle$ Protein x Days)		
Albumin	Globulin			Total Protein	Albumin	Globulin
gms %	gms %		gms/day/sq.M.			
2.4	2.0	Bovine serum	0	0	0	0
2.4	2.3	Bovine serum	+0.86	0.7	0.7	0
2.4	1.9	Bovine serum	+2.16	5.1	5.1	0
1.6	2.6	Bovine serum	+2.80	3.6	1.7	−1.1
2.5	2.0	Casein	+1.96	2.6	1.3	1.3
2.2	2.1	Casein	+3.18	3.3	1.4	1.9
2.2	2.4	Squibb hydrolysate	+2.78	4.6	1.5	3.1
2.0	2.1	Squibb hydrolysate	+3.02	3.4	1.6	1.8

the effects of nitrogen balances and protein source on regeneration areas, obtained under the experimental conditions used. The first two columns in this table describe the depleted state of the dog in terms of plasma albumin and globulin concentrations. The third column records the source of protein fed, and the fourth column, the nitrogen balance obtained by these feedings. The last three columns give the areas under curves illustrating changes in plasma protein concentration from those in the depleted state. The data on bovine serum protein illustrate, again, that the magnitude of regeneration of plasma proteins becomes greater as the nitrogen balance increases on the positive side. All of the regeneration, following the feeding of bovine serum protein, was restricted to the albumin fraction. On the other hand, casein and Squibb casein hydrolysate stimulated the regeneration of both the albumin and globulin fractions. Such differences in abilities of proteins to stimulate the formation of plasma proteins give impetus to the thought that different patterns of amino acids will have different jobs to do in protein therapy.

SUMMARY

The relation between nitrogen balance (*NB*) and either total nitrogen intake (*NI*) or absorbed nitrogen (*AN*) was demonstrated to be linear in the region of negative or low positive nitrogen balance, becoming curvilinear, well on the positive side of nitrogen balance. The rate of change of nitrogen balance, with respect to nitrogen intake, was called the nitrogen balance index. Thus, K and K' in the following equation are the nitrogen balance indices of absorbed nitrogen (AN) and total nitrogen intake (NI), respectively.

$$NB \quad K\ (AN) - NE_0 \text{ and} \tag{8}$$

$$NB \quad K'\ (NI) - NE_0, \tag{9}$$

where NE_0 is the excretion of nitrogen on a protein-free diet. K' divided by K is the digestibility of the protein, and K' multiplied by the fraction of protein in the protein source gives the "net protein" value.

The relationship between urine nitrogen (UN) excretion and nitrogen intake can be derived from EQUATIONS 8 and 9 to be:

$$UN = (1 - K)\ AN + UN_0, \tag{10}$$

where UN_0 is the excretion of urine nitrogen on a protein-free diet.

The slope of the line describing the relation between absorbed nitro-

gen and nitrogen balance is a function of the fraction of nitrogen retained in the body of the animal. It is this slope which will hereafter be referred to as the nitrogen balance index.

The nitrogen balance index of egg albumin was determined to be close to unity, in adult dogs receiving an adequate caloric intake. A reduction in caloric intake did not affect the utilization of egg albumin nitrogen, as measured by the nitrogen balance index, until the calories were reduced to less than 50 per cent of adequate. At 25 per cent of an adequate caloric intake, the utilization of egg white nitrogen was reduced markedly. The protein-sparing action of carbohydrates and of fats was illustrated in these studies on caloric intake. The nitrogen balance indices of some protein sources are recorded as follows: lactalbumin 1.0, dried egg albumin 0.96, casein 0.80, Squibb casein hydrolysate 0.80, Amigen 0.80, chicken entrails 0.77, flounder entrails 0.77, flounder heads 0.52, Protein A 0.32. The nitrogen balance index was shown to be constant, while protein minima for nitrogen equilibrium varied.

Similar relationships between nitrogen input and output were found in hypoproteinemic, as in control, dogs. The reduction in protein stores in hypoproteinemic dogs was reflected by reductions below control values of excretion of nitrogen (particularly the urea and ammonia fractions) on a protein-free diet, and of protein minima for nitrogen equilibrium. The nitrogen balance index increased above control values in the hypoproteinemic dogs, proving greater retention of nitrogen by protein-depleted, than by normal, animals. Nitrogen balance indices greater than unity have been obtained while feeding egg albumin nitrogen to hypoproteinemic dogs. Thus, egg white nitrogen spares body nitrogen in these protein-depleted dogs.

The relationship between the fraction of nitrogen retained in the body of the animal (BV) and the nitrogen balance index (K), and so-called endogenous nitrogen, was illustrated by EQUATION 11:

$$EN = NE_0 - AN\ (K - BV) \qquad (11)$$

If K is equal to BV, then EN is equal to NE_0, but, if K is greater than BV, then the excretion of endogenous nitrogen is less than NE_0, and body nitrogen is spared.

The excretion of nitrogen was reduced below NE_0, in normal dogs, when methionine was added to the protein-free diet or to diets containing casein or egg albumin. Thus, the nitrogen-sparing action of methionine was demonstrated.

A constant state of depletion was produced in the hypoproteinemic dogs, after the total plasma proteins were reduced by plasmapheresis and protein-free feeding to approximately 4 grams per cent. The plasma protein did not always increase the first day or two following feeding of nitrogen to these dogs, even though they were well on the positive side of nitrogen balance. After this lag period, the plasma protein increased rapidly, for a few days, followed by a slower rise, so that normal plasma protein concentration was reached in about three to four weeks of feeding. Plasma volumes usually started to increase about the fifth day of nitrogen feeding. Since the most rapid regeneration usually took place within the first five days of feeding of nitrogen, without changes in plasma volume, the following technique was developed for rapid assays.

The dogs were depleted by feeding the protein-free diet and by a few days of plasmapheresis, until the plasma protein concentration was between 4.0 and 4.5 gms. per 100 ml. plasma and essentially constant, under the experimental conditions used. The depleted animals were fed protein nitrogen for five days, and then returned to the protein-free diet until the plasma protein concentration had returned to the former depleted level. Plasma protein, in gms. per cent, was plotted against days, and the area under the curves showing the increase above the depleted concentration was taken as a measure of the amount of regeneration. These areas ($\triangle$ protein $\times$ days) take into consideration, not only the magnitude of the rise in plasma protein, but also the rate of depletion following nitrogen feeding, giving better correlation with nitrogen balance than could be obtained with regeneration of plasma proteins alone.

Regeneration of plasma proteins, measured by these areas, increased in magnitude as positive nitrogen balances increased. Thus, regeneration of plasma proteins in hypoproteinemic dogs is a function of the nitrogen balance produced and, therefore, of the nitrogen balance index. The kind of plasma protein regenerated, however, depended, in part, upon the pattern of amino acids fed. Bovine serum protein, for example, promoted regeneration of serum albumin, while casein stimulated the formation of both albumin and globulin fractions.

BIBLIOGRAPHY

1. **Allison, J. B., & J. A. Anderson**
1945. J. Nutrition **29**: 413–420.

2. **Allison, J. B., R. D. Seeley, J. H. Brown, & J. A. Anderson**
1946. J. Nutrition **30:**
2a. **Allison, J. B., J. A. Anderson, & R. D. Seeley**
1946. (In Press.)
3. **Bricker, M., H. H. Mitchell, & G. M. Kinsman**
1945. J. Nutrition **30:** 269–283.
4. **Chow, B. F., J. B. Allison, W. H. Cole, & R. D. Seeley**
1945. Proc. Soc. Exp. Biol. & Med. **60:** 14–17.
5. **Cowgill, G. R.**
1928. Am. J. Physiol. **85:** 45.
6. **Elman, R., & H. W. Davey**
1943. J. Exp. Med. **77:** 1.
7. **Gregersen, M. I., & J. D. Stewart**
1939. Am. J. Physiol. **125:** 142.
8. **Holman, R. L., E. B. Mahoney, & G. H. Whipple**
1934. J. Exp. Med. **59:** 251.
9. **Howe, P. E.**
1921. J. Biol. Chem. **49:** 109.
10. **Madden, S. C., & G. H. Whipple**
1940. Physiol. Rev. **20:** 194.
11. **Melnick, D., & G. R. Cowgill**
1937. J. Nutrition **13:** 401.
12. **Melnick, D., G. R. Cowgill, & E. Burack**
1936. J. Exp. Med. **64:** 897.
13. **Miller, L. L.**
1944. J. Biol. Chem. **152:** 603.
14. **Mitchell, H. H.**
1944. Ind. Eng. Chem., An. Ed. **16:** 696.
15. **Mitchell, H. H., & G. G. Carman**
1924. J. Biol. Chem. **60:** 613.
16. **Robinson, H. W., J. W. Price, & C. G. Hogden**
1937. J. Biol. Chem. **120:** 481.
17. **Schoenheimer, R., & D. Rittenberg**
1940. Physiol. Revs. **20:** 218.
18. **Seeley, R. D.**
1945. Am. J. Physiol. **144:** 369.
19. **Weech, A. A.**
1938–1939. The Harvey Lectures **34:** 57.
20. **Weech, A. A., & E. Goettsch**
1938. Bull. Johns Hopkins Hosp. **53:** 154.
21. **Weech, A. A., E. Goettsch, & E. B. Reeves**
1935. J. Exp. Med. **61:** 299.
22. **Willman, W., P. P. Swanson, G. F. Stewart, G. T. Stevenson, & M. Brush**
1945. Proc. Fed. Am. Soc. Exp. Biol. **4:** 164.

DISCUSSION OF THE PAPER

Dr. D. V. Frost (*Abbott Laboratories, North Chicago, Illinois*):

The work of the Rutgers group, as presented in this paper by Doctor Allison, has brought something new to the classical concept of nitrogen balance and biological value and has done much to clarify our thinking in the use of these terms. The graphic method should find considerable application in studies with various species, and offers a stimulating new approach to certain complex problems. It provides a new key, for instance, to estimation of the so-called endogenous nitrogen excretion, which has long been a point of question in the conventional method of determining biological value.

Certain difficulties arise in making the nitrogen balance method of maximum value in assessing protein efficiency. Doctor Allison has indicated what some of these difficulties are, and I would like to raise one or two further questions for his consideration and comment.

Would it be well, if dogs were conditioned on a standard protein source, at a standard nitrogen level, prior to any experiment? Such procedure may avoid the error inherent in the fact that the hypoproteinemic animal shows greater avidity for almost all proteins than does the normal animal. Questions of the nutritive state would also be partly resolved by such standardization, so far as later interpretation is concerned.

Such factors as the sparing action of methionine and cystine for nitrogen excretion, and the somewhat similar sparing action of proteins of high biologic value for body nitrogen, are difficult to reconcile in any scheme for devising good and reliable "endogenous nitrogen" values. However, good, representative values, which will fit a strict empirical definition, may be devised and be of real value. Since the nitrogen loss on nitrogen-free diet decreases gradually, with loss in body nitrogen stores, it is difficult to achieve constancy by this method. The better value in the normal dog would appear to be the nitrogen loss, on a minimum level of a completely digested protein. Would it be well to try to establish the use of a pure, standard protein for this purpose and, going a step further, attempt to establish figures for endogenous nitrogen loss, on a kilogram or square meter body surface basis, which would be generally useful? These values may be correlated with figures obtained by the graphic method described by Doctor Allison. All of this presumes a high degree of standardization in methods, but would appear to be the only recourse, where empiricism exists and where all of the variables may not yet be known.

I would like, also, to present some work of our own, in line with the work of the Rutgers group. We became interested in applying Doctor Allison's graphic method of calculating biological value to data which we had at hand on intravenous utilization of partial acid hydrolysates, prepared from beef blood fibrin (Armour). Dr. W. C. Risser and I were interested in determining the minimum level of a partial acid hydrolysate of fibrin needed to give nitrogen balance in dogs, when given by slow intravenous injection. The dogs used were somewhat hypoproteinemic, having been on submimmal nitrogen levels for many months. A start was made at 85 mg. N per kilo per day, at which level nitrogen balance was negative. Increases were made gradually, over a period of six weeks, to a level of 110 mg. N, where the dogs came into positive balance. The study was continued for a total of 15 weeks, during the latter part of which there was slight positive balance. Weight loss occurred gradually, throughout the study, although the dogs ate the low protein diet, at a rate of 60 per cent of the 80 calories per kilo per day offered, or an average rate of 48 calories per kilo per day.

A graph of nitrogen balance against nitrogen injected (absorbed nitrogen) gave a series of 14 points for each of two dogs. Representative lines drawn through the points had slopes of .83 and .86 and are the nitrogen balance indices, according to the formula $NB = BV(AN) - EN$. It is, perhaps, significant to note that our values were expressed in terms of mg. nitrogen per kilo instead of Gm. nitrogen per square meter body surface, as used by Doctor Allison. We used the changing weights of the dogs, rather than the standard weights, and Doctor Allison informs me this is also done in making their calculation of body surface. A point of interest in our study is the fact that weight loss continued at a slow rate, even after the nitrogen balance had become definitely positive. A possible interpretation of this phenomenon is that loss of body fluid may occur, as protein stores are replenished. This effect would be the converse of the increase in body fluids which is associated with the hypoproteinemic state.

The best approach to minimum levels which will give nitrogen balance may also be open to question. Melnick and Cowgill[1] have concluded, on logical grounds, that the most workable approach is from the side of negative balance. If this is done, however, as in the case of our animals, weight balance, even in the face of slight positive nitrogen balance, may not be maintained, when the experiment

[1] **Melnick, D., & G. R. Cowgill.** J. Nutrition 13: 401. 1937.

continues over a long period. It may be suspected that, in such a condition, there is some unusual realignment of body protein stores, to give maximal conservation of certain key amino acids. In the long run, this may lead to ill effects and final breakdown. Certainly, it is true that the well-fed animal requires much more nitrogen from a given source, to maintain both nitrogen and weight equilibrium, than does the hypoproteinemic animal. The well animal remains in obvious health and in buoyant spirits and, in doing so, appears to be quite prodigal of food nitrogen. The hypoproteinemic animal, on the other hand, may come into bare nitrogen equilibrium or slight positive balance at low levels of nitrogen, but remains apathetic and may even continue to lose weight. It would seem, therefore, that such additional criteria as weight balance, the level of total circulating plasma protein, and the general health of the animal must be considered, together with nitrogen balance data, when minimum protein requirements are being determined.

Dr. Daniel Melnick (*Food Research Laboratories, Inc., Long Island City, N. Y.*):

The paper by Drs. Allison, Anderson, and Seeley constitutes such a comprehensive coverage of the subject of nitrogen balance that very little in the way of supplementary remarks is in order. We had found in studies with lactalbumin, serum protein, casein, and gliadin (proteins which vary widely in biological value), that a linear relationship existed between the nitrogen balance and nitrogen intake, in the region of negative and low positive nitrogen balance. The intercept of the straight lines, showing this relationship on the abscissa representing nitrogen equilibrium, gave the protein minima, expressed in terms of the per cent of protein calories in the ration. Dr. Allison and associates have greatly improved the assay technic, by first standardizing the dogs with a protein-free ration. This allowed estimations of the nitrogen balance index of the test proteins. This represents the rate of change of nitrogen balance, with respect to absorbed nitrogen, *i.e.*, the slope of the response curve.

Of interest is the finding, by Dr. Allison and his group, that there is a critical intake of calories, below which utilization of protein (*i.e.*, retention of nitrogen) is impaired. It is true, and I say this on the basis of nitrogen retention studies conducted on the rat, that increasing the protein content of the ration leads to progressively poorer utilization of nitrogen for tissue anabolism, a greater portion of the protein in the ration serving merely as a source of calories. In Dr. Allison's study, however, restriction of calories from non-protein sources, without compensatory increase in protein calories, leaves the organism a difficult choice: to use the protein for tissue synthesis, or for caloric purposes. Apparently, only when the restriction in total caloric intake is markedly severe, will the organism fail to use dietary protein for the synthesis of tissue protein.

It may be of interest to this group to mention that Mitchell and his associates, at the University of Illinois, recently reported the results of studies on protein minima for nitrogen equilibrium, using natural foods as the source of protein, and humans as the test subjects. Essentially, the same technic as described by Dr. Allison was employed. They, too, observed that the individual test subjects may differ as to their protein minimum, but that they tend to show a parallel response to the various levels of dietary protein intake. This finding, reported, now, from three different laboratories, should find application in studies with hydrolysates and pure amino acid mixtures concerned with establishing their minima for nitrogen equilibrium. Such dietary mixtures may be either too costly or unpalatable to allow serial studies on a large group of subjects. It is suggested that two or three individuals be employed to establish the slope of the curve relating response to level of nitrogen intake. For each of the remaining subjects, it should suffice to conduct only one nitrogen balance study and extrapolate to the line of nitrogen equilibrium, by drawing a line parallel to the average curve, established with the first group of individuals. The nitrogen balances for the supplementary subjects should, in each case, be slightly negative, thereby reducing to a minimum any error due to assuming the slope of the response curve for each individual to be the same as for the mean.

Returning to Dr. Allison's report, I would like to comment on one other phase. He states: "The altered physiological state of the hypoproteinemic dog increases

the utilization of protein nitrogen." This statement needs some qualification. Dr. Allison's experiments actually demonstrate that the altered physiological state of the dog, subsisting on a protein intake so inadequate that a reduction in the plasma protein concentration results, is responsible for increased utilization of protein nitrogen, during the period of re-alimentation. The reports of others, cited by Dr. Allison, demonstrate that, during depletion by feeding a protein-free diet, the loss of plasma protein constitutes only a small fraction (about 4 per cent) of the total loss of nitrogen. In Dr. Allison's study, a condition of hypoproteinemia, dissociated from general protein depletion, was not produced. Indeed, it was for this reason we had criticized reports claiming to be evaluations of proteins for plasma protein regeneration, when based upon the prolonged feeding of protein-inadequate diets for the production of hypoproteinemia. It may be desirable to study the recovery of such depleted animals and rate proteins according to the responses elicited. However, it is improper to attribute this same rating to the proteins, in so far as the specific function (if one exists) of plasma protein formation is concerned.

In our studies on the influence of diet on serum protein regeneration, attempts were made to correct for that moiety of protein intake required for nitrogen equilibrium, using the plasmapheresis technic for the production and maintenance of the hypoproteinemia. Under such circumstances, it was found that the three more or less complete proteins, serum protein, lactalbumin, and casein, were approximately equal, in promoting the regeneration of serum protein.

THE EVALUATION OF PROTEIN QUALITY IN THE NORMAL ANIMAL

By Richard H. Barnes and David K. Bosshardt

From the Department of Biochemical Research, Sharp and Dohme, Inc., Glenolden, Pennsylvania

There are available to the investigator several methods of measuring the nutritive adequacy of protein foods. These methods provide a good classification of dietary proteins, and their wide application has resulted in a rather comprehensive knowledge of many single protein sources and protein mixtures. The methods are more or less empirical, and occasionally values are obtained that are sufficiently out of line to lead one to suspect that a lack of adequate control was employed. It must be emphasized that these methods, empirical as they may be, have proven to be of considerable practical value. It is seldom that one observes a serious reversal in the relative classification of proteins, although variations in the absolute values obtained by any one method may be large. The practical usefulness of these methods, in their present form, should not cause a stagnation of effort toward improvements and a more thorough understanding of metabolic phenomena that are involved, for these are pathways that will lead to a more complete knowledge of protein nutrition.

In this discussion, it is intended to present a few of the more or less controversial factors that are evident in the measurements of protein quality, for growth or maintenance in the normal animal. The factors that will be discussed are the following:

A. Growth

1. Growth rate, as an index of nutritive quality.
2. The influence of the level of protein in the diet on the utilization of protein for growth.
3. Paired-feeding, versus *ad libitum* feeding.
4. The influence of caloric intake on the utilization of protein for growth.
5. Body weight gain, as an index of body protein gain.

B. Maintenance

6. The measurement of "metabolic" fecal nitrogen.
7. The measurement of "endogenous" urinary nitrogen.
8. The measurement of the combined utilization of protein for growth and maintenance.

No attempt will be made at completeness, in the discussion of these subjects. Rather, each subject will be presented in the light of studies that have been conducted in this, and closely associated, laboratories.

A. GROWTH

Growth Curves as an Index of Nutritive Quality

A very practical method of expressing the nutritive quality of proteins would be in terms of the amount of protein necessary for the

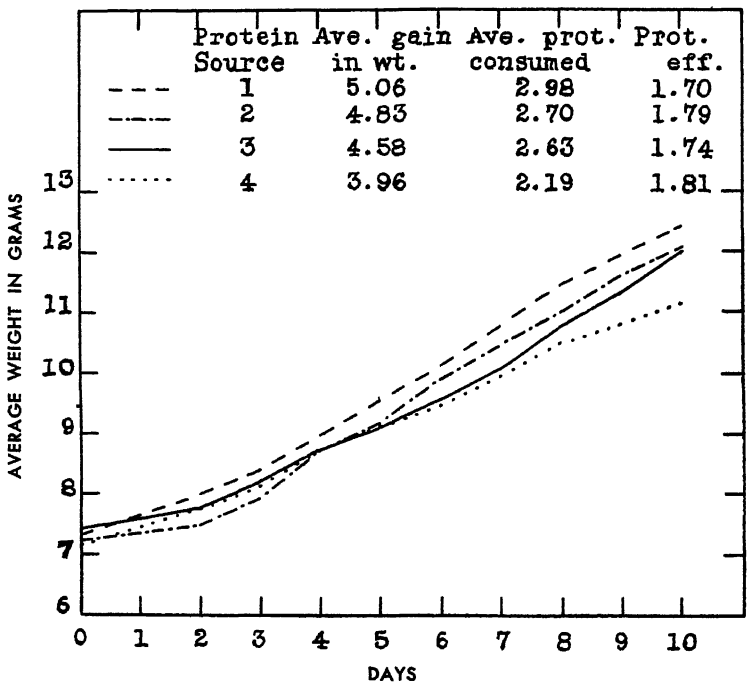


FIGURE 1. Mouse growth rates, illustrating a relative constancy of protein efficiency values (grams gain in body weight per gram of protein consumed), with changing growth rates.

attainment of a certain growth rate. The classification of proteins or protein mixtures, on this basis, would provide a direct index of the

amount of any protein that would be necessary, in order to provide for a given growth rate. For several reasons, this method of classification is not feasible for the analysis of a large group of protein sources. Furthermore, the method does not provide information concerning the fundamental nutritive characteristics of the protein *per se*, but rather gives an index of the nutritive quality of the entire diet.

Examples showing the lack of relationship that may exist between the nutritive quality of proteins and growth rate may be seen in FIGURES 1 and 2. The four proteins, in FIGURE 1, have the same nutritive quality, on the basis of Osborne, Mendel, and Ferry's[1] protein efficiency measurement, but growth rates of the mice that received these proteins are not the same. In FIGURE 2, the reverse situation is found. In this case,

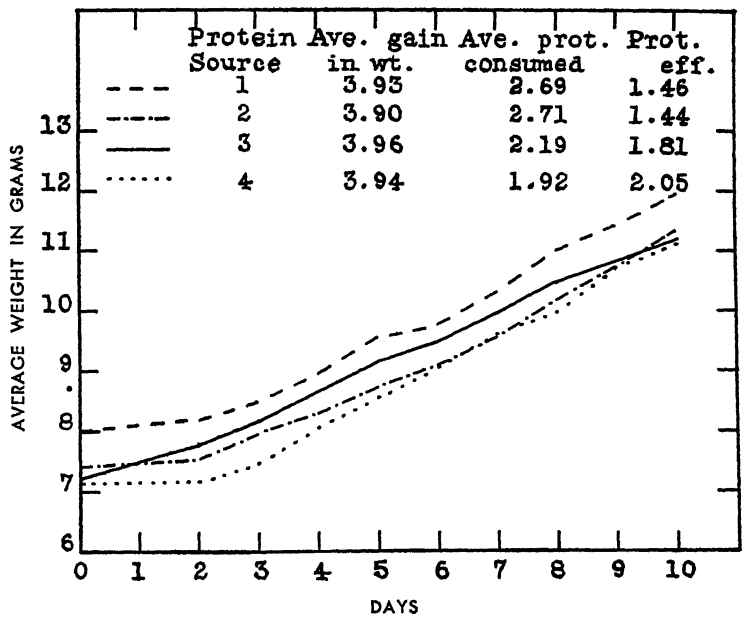


FIGURE 2. Mouse growth rates, illustrating a relative constancy of growth rates, with wide variations in protein efficiency values.

mice growing at the same rate are found to have different protein efficiencies. The inadequacy of this method of expressing protein quality is recognized by many nutritionists, and these data are presented merely as confirmatory evidence. In spite of the practicality of expressing protein quality on the basis of growth rate, it would appear that this method is not adequate for the true evaluation of nutritive quality.

Influence of the Level of Protein in the Diet on the Utilization of Protein for Growth

Osborne, Mendel, and their associates[1, 2] were the first to show the influence of the protein content of the diet on protein utilization, as measured by the gain in weight, resulting from the ingestion of a unit weight of protein. Little recognition has been given to the importance

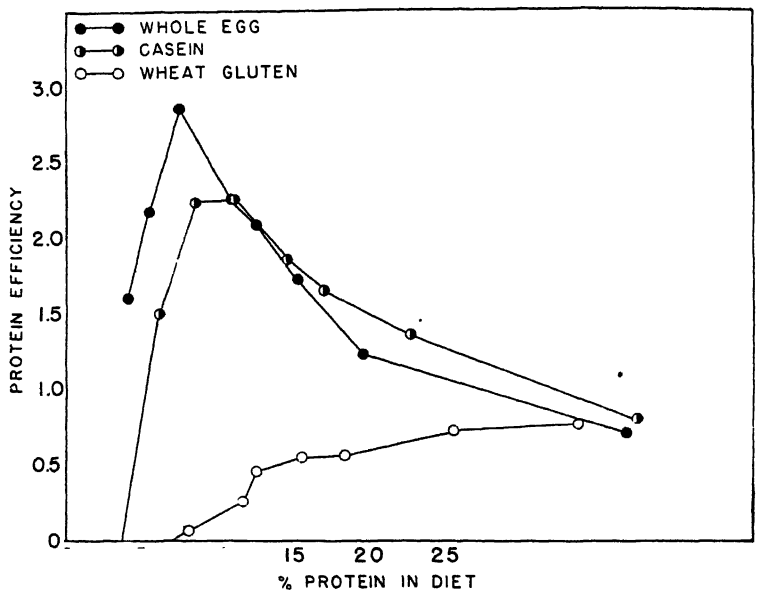


FIGURE 3. The influence of the level of protein in the diet on protein efficiency values, as measured with mice. Protein efficiency is expressed as grams gain in weight per gram of protein consumed, and is calculated for 10 day feeding periods. (**Bosshardt, D. K., L. C. Ydse, M. M. Ayres, & R. H. Barnes.** J. Nutrition **31**: 23. 1946.)

of the level of protein in the diet, on the value of the protein efficiency ratio. As a result, this widely employed method is frequently misused. The magnitude of the change in the protein efficiency ratio, resulting from the incorporation of different amounts of protein in the diet, is illustrated in FIGURES 3 and 4. The originators of this method stipulated that the maximum ratio should be established, and this ratio then taken as the index of protein quality for growth. The common practice of employing a 10% protein diet, regardless of the nature of the protein, will result in a considerable distortion of nutritive values, and the magnitude of the error will increase as the nutritive quality of the protein decreases.

Paired-Feeding *versus ad libitum* Feeding

It is difficult to rationalize any method for measuring the nutritive quality of protein, if the amounts ingested of all proteins studied are not equal. This point has been emphasized by Mitchell and his associates,[3] and has resulted in the wide application of the so-called paired-feeding method. In TABLE 1, is illustrated the amount of restriction

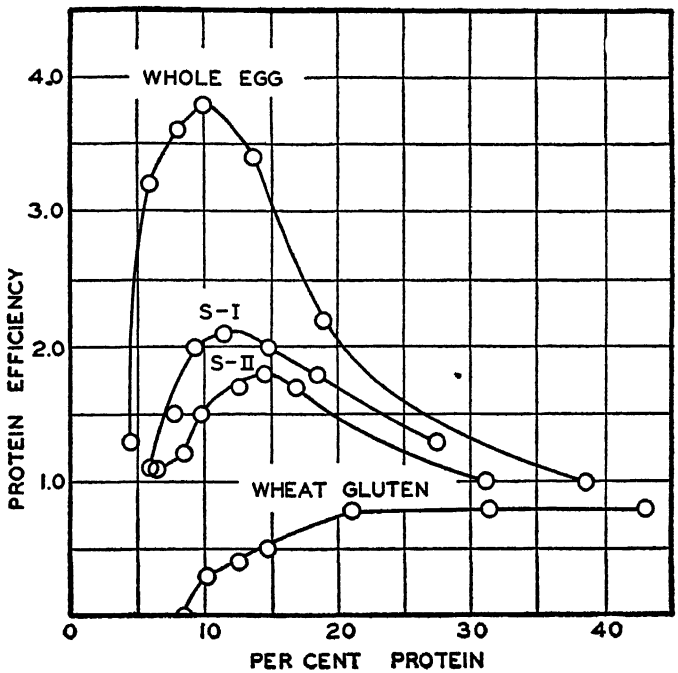


FIGURE 4. The influence of the level of protein in the diet on protein efficiency values, as measured with rats. Calculations are for 42 day feeding periods. (**Barnes, R. H., J. E. Maack, M. J. Knights, & G. O. Burr.** Cereal Chemistry. **22**: 273. 1945.)

that may be imposed, when the two protein sources vary widely in nutritive quality. In both examples, typical paired-feeding experiments were carried out, over a 54 day period. At the end of this period, *ad libitum* feeding was begun. This had no effect on the poorer protein groups, because they were already eating to the limit of their appetites. The restricted pair-mates, receiving the better protein, immediately increased their food consumption and growth rate. In the first case, whole wheat flour was compared with a heat-treated wheat of distinctly poorer protein quality, and, in the second case,

ground white rice was compared with a heat-treated rice, also of distinctly poorer protein quality. It is clear that paired-feeding, as practised by many, gives an entirely artificial value for the better protein. If the restriction is severe, some of the better protein may be wasted

TABLE 1

COMPARATIVE RESULTS WITH PAIRED AND *ad libitum* FEEDING, WHEN THE POORER DIET IS VERY RESTRICTING

The whole wheat diet contained 11.2% protein, and the white rice diet, 5.65% protein, expressed on a dry matter basis (Barnes, R. H., J. E. Maack, M. J. Knights, & G. O. Burr. Cereal Chemistry, 22: 273. 1945).

	Average food intake g/day	Average gain in weight g/day	Average protein efficiency weight gain per gram protein	Ratio of P.E. by *ad libitum* to P.E. by restricted feeding
Whole wheat diet restricted by paired-feeding	3 44	0 157	0 42	2 43
Whole wheat diet following release to *ad libitum* feeding	7 46	0 811	1 03	
White rice diet restricted by paired-feeding	4 77	0 228	0 92	1 87
White rice diet following release to *ad libitum* feeding	7 03	0 711	1 72	

as fuel, for which sucrose or lard could have served as well. A more complete discussion of the influence of caloric restriction will be considered later. In TABLE 2, is presented a comparison of protein efficiency values, obtained at the maximum by *ad libitum* feeding; at a 10% protein level, by *ad libitum* feeding; and at a 10% protein level,

TABLE 2

EFFECT OF EQUALIZING THE PROTEIN INTAKE ON PROTEIN EFFICIENCY VALUES

(Barnes, R. H., J. E. Maack, M. J. Knights, & G. O. Burr. Cereal Chemistry. 22: 273. 1945.)

Protein source	Maximum P.E. *ad libitum* feeding	P.E. at 10% protein *ad libitum* feeding	P.E. at 10% protein pair-fed
Wheat gluten	0.8	0.3	0.4
Soyflour No. 2	1.8	1.5	1.2
Soyflour No. 1	2.1	2.0	1.8
Whole egg	3.8	3.8	2.6

by paired feeding. Although considerable differences in values are evident, these data alone do not permit one to conclude which is the more valid method of expressing nutritive quality.

FIGURES 5 and 6 show the percentage utilization of absorbed protein for body protein gain, in the mouse and rat. In each case, the protein source was incorporated into an iso-caloric diet, at several levels. In

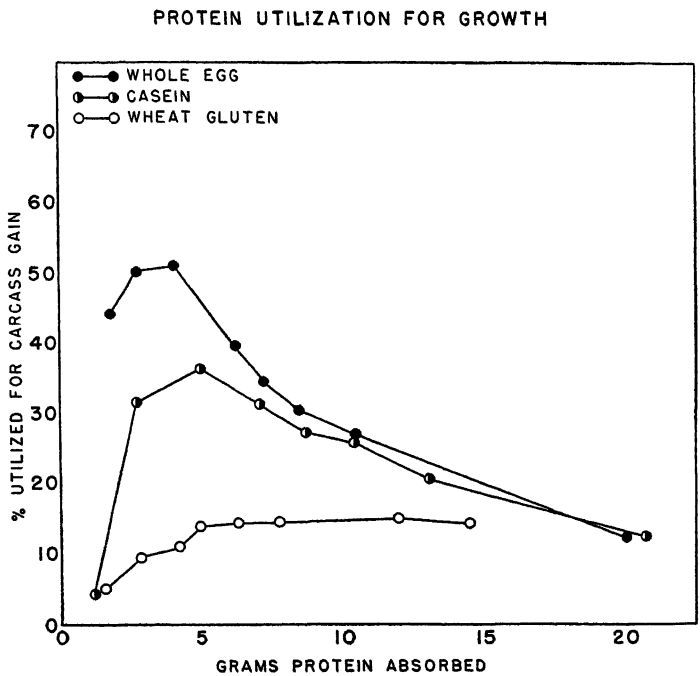


FIGURE 5 Protein utilization for body protein gain in the mouse Body protein gain is for 20 day feeding periods. Protein absorbed is corrected for metabolic fecal nitrogen. (**Bosshardt, D. K., L. C. Ydse, M. M. Ayres, & R. H. Barnes.** J. Nutrition. **31**: 23. 1946.)

the rat studies, the animals were killed after 42 days and, in the mouse studies, after 20 days. Total nitrogen was determined on the carcasses minus intestinal contents, at the termination of the experiment, and body nitrogen, at the start of the experiment, was measured on control groups of animals. These data show that protein utilization for growth reaches maximal values at essentially the same level of absorbed protein. There is some slight deviation of the maximal points, but, even with proteins of widely differing nutritive quality, this deviation does not introduce any appreciable change in the calculated percentage utilization for growth. For practical purposes, it may be said

that, at maximal protein utilization, absorbed protein is the same. This illustrates how it is possible to equalize protein intake, by *ad libitum* feeding, and, therefore, to obtain the maximal protein utilization, without introducing an artificial food restriction.

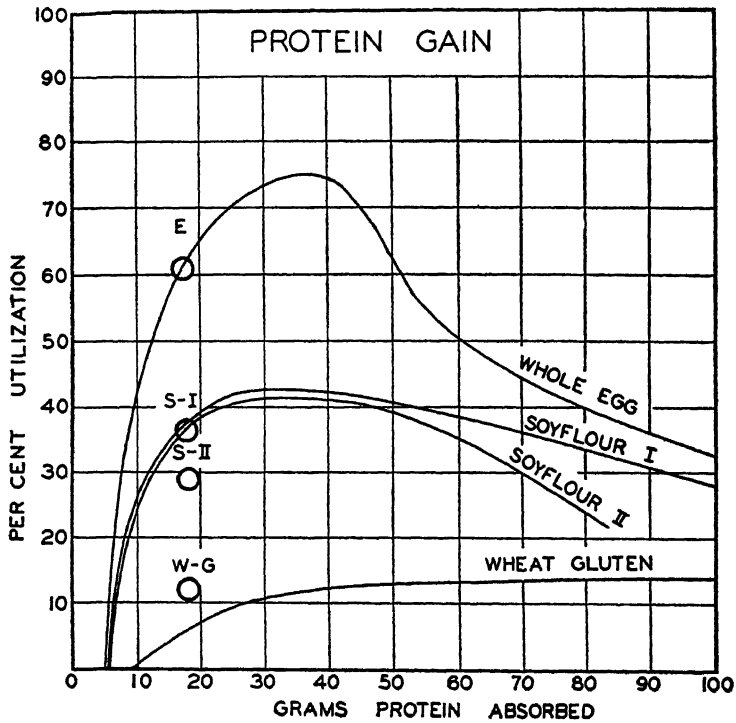


FIGURE 6. Protein utilization for body protein gain in the rat. Body protein gain is for 42 day feeding periods. Protein absorbed is not corrected for metabolic fecal nitrogen. (**Barnes, R. H., J. B. Maack, M. J. Knights, & G. O. Burr.** Cereal Chemistry. **22**: 273. 1945.)

In FIGURE 6, the large circles represent protein utilization in pair-fed animals. It will be seen that, at this level of protein intake, protein utilization is on the ascending slope, but these data do not indicate any definite wastage of protein due to the caloric restriction that is imposed by reducing food intake to the level of the poorest diet that is consumed. However, it must be kept in mind that, even with *ad libitum* feeding, a caloric restriction may be imposed, due to the decreased appetite resulting from the incorporation of very small amounts of protein in the diet.

The Influence of Caloric Intake on the Utilization of Protein for Growth

In the adult animal, a restriction of caloric intake results in a decrease in the utilization of protein. This has been related to the well-known sparing action of carbohydrate on urinary nitrogen excretion,[4] but the data of Allison and his associates[5] indicate that both fat and carbohydrate calories are effective in changing protein utilization. It is of considerable interest that these investigators find that protein utilization, in the adult dog, is not noticeably decreased, until the caloric intake is dropped to approximately 25 per cent of its normal maintenance level. Larson and Chaikoff[6] have noted that the nitrogen excretion of dogs receiving a "normal," purified diet may be decreased, by supplying extra carbohydrate. In particular, they noted that the sparing action of the extra carbohydrate was dependent upon the time interval between feeding the normal, protein-containing meal and the extra carbohydrate.

In the growing rat, the caloric decrease that results from pair-feeding a poor protein together with a good protein, may have some influence upon protein utilization. This is frequently cited as a criticism against the paired-feeding technique. However, in the paired-feeding studies mentioned earlier (FIGURE 6), there was no obvious decrease in protein utilization due to food restriction *per se*. This should not be considered as complete evidence that a small caloric restriction has no influence upon protein utilization, for the *ad libitum* animals that are used for comparison may also have a caloric restriction.

TABLE 3

AVERAGE CALORIC INTAKE

Per 100 square centimeters body surface area (weight ⅔. 11.36) of rats consuming diets containing whole egg protein. Food consumption was *ad libitum* and averages are for 42 day feeding periods.

Diet No.	Protein calories in diet	Average B.S. area	Calories per 100 cm² B.S. (average—42 days)
	%	Cm²	
1	3.7	173	11.1
2	5.3	212	14.7
3	7.2	236	15.0
4	9.1	281	15.3
5	12.5	302	14.9
6	17.4	306	14.2
7	35.4	308	13.8

☐ Enclosed figures are at maximal protein utilization

In TABLE 3, the average daily energy intake of growing rats, per 100 square centimeters body surface area, is shown. These groups received diets containing whole egg protein for 42 days, and were allowed to eat *ad libitum*. It will be seen that, at the level of maximal protein utilization, there was maximal consumption of total calories, for the attainment of a given body size. Caloric intake, per unit body surface area, decreases on either side of this point. Another example is presented in TABLE 4. Mice, consuming iso-caloric diets containing

TABLE 4

AVERAGE CALORIC INTAKE

On the basis of weight[2], of mice consuming diets containing casein. Food consumption was *ad libitum* and averages are for 20 day feeding periods.

Diet No.	Protein calories in diet	Average weight[2]	Calories per gm. ave. wt.[2]
	%	gm.	
1	2.79	3.81	2.34
2	4.90	4.52	2.52
3	6.67	5.33	2.87
4	8.72	5.86	2.86
5	11.52	5.84	2.67
6	13.30	6.01	2.69
7	17.80	6.07	2.51
8	29.40	5.93	2.46

☐ Enclosed figures are at maximal protein utilization.

different levels of protein, exhibit a similar relationship between caloric intake and protein utilization. This same type of calculation has been made for rats receiving 4 different proteins, and for mice receiving 3 different proteins. In each of the 7 cases, the pattern of caloric utilization was the same. Although it cannot be concluded that optimal calories are supplied, when protein utilization is maximal, these data do show that the greatest caloric consumption, per unit body surface area, occurs at the level of maximal protein utilization. The lower caloric consumption, on either side of maximal protein utilization, may have a marked effect on the extent of protein utilization.

In TABLE 5, is shown caloric consumption, as a result of the *ad libitum* feeding of a series of diets containing a well heat-treated soyflour. In all other experiments, diets were made equicaloric by replacing sucrose with protein. In this series, there are progressive changes in protein content, but the caloric values of the diets, also, are altered by chang-

ing the fat and cellulose, as well as sucrose and protein content. The diets vary in caloric content, from approximately 2.5 to 6.6 calories per gram. In one set (group 1), the diet with the lowest protein content also has the lowest caloric value, and in the other (group 7), the lowest protein content is in the highest caloric diet. It is intended to discuss

TABLE 5

CALORIC INTAKE

As a function of surface area, for rats receiving diets containing different amounts of soyflour protein for 42 days. The caloric value of the diets was varied by changing the fat and fiber content.

Diet No.	Protein weight in diet	Protein calories in diet	Caloric value at diet	Protein consumed	Average B.S. area	Calories per 100 cm^2 B.S. (average—42 days)
	%	%	Cal.	gm.	Cm^2	
1	6.8	11.1	2.50	27.2	195	12.2
2	7.6	9.5	3.30	26.6	220	12.6
3	10.4	10.0	4.25	37.0	233	15.4
4	11.4	9.5	4.93	35.1	238	15.2
5	14.3	10.2	5.73	46.5	259	17.2
6	15.7	9.8	6.58	45.8	262	17.4
7	9.1	5.7	6.52	22.0	210	17.9
8	10.4	8.8	4.83	33.5	235	15.8
9	10.0	9.6	4.27	35.8	236	15.4
10	9.8	11.5	3.51	44.4	249	15.2
11	10.1	13.8	3.01	49.0	249	14.0
12	10.7	16.7	2.63	58.0	260	13.0

☐ Enclosed figures are at maximal protein utilization

the results of this study, from the standpoint of protein utilization alone. There are many interesting points, with regard to voluntary food intake, that have arisen from this study, but these factors do not enter into the scope of this presentation. It will be noted that, in the first 6 groups, the highest ratio of calories to body surface area is at the level of highest protein intake, but, in the second series, the highest caloric ratio occurs at the lowest protein intake. With one exception, diet 7, the highest caloric intake is associated with the maximal protein efficiencies. This same protein source has also been fed, in iso-caloric diets, to groups of rats. It is of interest to compare caloric intake with protein utilization, at different levels of protein intake, in the 2 experiments. The left section of TABLE 6 shows caloric consumption

and protein utilization, at the various levels of protein intake for rats that received the diets having a wide range in caloric value. These are the same data that were shown in TABLE 5. The right section shows interpolated values of caloric consumption and protein efficiency ratios, at the same levels of protein intake, but, in this case, the rats received iso-caloric diets of varying protein content. Comparisons must be made at the same level of protein intake, for it has been shown that

TABLE 6

THE RELATIONSHIPS BETWEEN CALORIC INTAKE PER UNIT SURFACE AREA AND PROTEIN EFFICIENCY VALUES AT EQUAL LEVELS OF PROTEIN INTAKE

The protein source employed in both sets was a well-heated soyflour.

Variation in caloric value of diet					Constant caloric value of diet*		
Diet No.	Caloric value of diet	Protein consumed	Calories 100 cm² B.S.	P.E.**	Caloric value of diet	Calories 100 cm² B.S.	P.E.**
1	2.50	27.2	12.2	1.3	4.5	16.4	1.5
2	3.30	26.6	12.6	1.9	4.5	16.4	1.5
3	4.25	37.0	15.4	1.9	4.5	15.7	2.0
4	4.93	35.1	15.2	1.9	4.5	15.4	2.0
5	5.73	46.5	17.2	2.2	4.5	15.4	2.0
6	6.58	45.8	17.4	2.4	4.5	15.4	2.0
7	6.52	22.0	17.9	1.9	4.5	16.7	1.5
8	4.83	33.5	15.8	1.9	4.5	16.0	2.1
9	4.27	35.8	15.4	2.1	4.5	15.4	2.0
10	3.51	44.4	15.2	1.7	4.5	15.4	2.0
11	3.01	49.0	14.0	1.7	4.5	15.5	2.0
12	2.63	58.0	13.0	1.7	4.5	15.7	2.0

* Some of these values were interpolated from curves.
** Protein efficiency; grams gain in weight per grams of protein consumed.

this is the primary factor governing the protein efficiency ratio. With one exception (diet 2), every case in which the protein efficiency ratio is higher, at a given level of protein intake, the caloric consumption per unit body surface area, is also higher. This experiment cannot be considered final proof that protein utilization in growing animals may be influenced by caloric intake, in spite of the fact that the animals consumed their food *ad libitum*. Protein utilization should be expressed in terms of body protein gain, rather than body weight gain, and some of the high protein efficiency ratios may be due to excess fat deposition, rather than protein. However, there is a strong indication that the utilization of protein is influenced by the caloric intake, under these conditions.

The conclusions drawn from this study are that, in a series of iso-

caloric diets, containing different amounts of protein and eaten *ad libitum*, the greatest caloric consumption, per unit body surface area, coincides with the level of protein intake that provides maximal protein utilization. However, there is some evidence that caloric intake is not optimal, even at the level of maximal protein utilization. The lower protein utilization that is found at protein intakes on either side of the maximal protein efficiency ratio are probably due, in part, to insufficient caloric intake.

Body Weight Gain as an Index of Body Protein Gain

A common argument against the expression of protein values in terms of body weight gain is that the increase in body weight may not give a true index of the increase in body protein.[7] Mitchell[8] and others have presented data showing that, in pair-fed rats, the nitrogen content of the carcass may differ, when the nutritive quality of the protein being fed is not the same. In TABLES 7 and 8, carcass protein (carcass

TABLE 7

PER CENT PROTEIN ($N \times 6.25$) IN RAT CARCASSES (MINUS INTESTINAL CONTENTS) AFTER RECEIVING EXPERIMENTAL DIETS FOR 42 DAYS

(**Barnes, R. H., J. E. Maack, M. J. Knights, & G. O. Burr.** Cereal Chemistry. 22: 273. 1945.)

Protein in diet (approximate)	Whole egg	Soyflour No. 1	Soyflour No. 2	Wheat gluten
%	%	%	%	%
		Ad libitum		
4	15.8	—	—	—
6	16.2	17.6	17.9	—
8	16.3	17.8	18.2	17.1
10	18.0	17.5	18.0	17.3
12	—	17.6	18.2	17.5
14	17.6	18.0	17.7	17.2
18	—	18.4	19.1	—
20	18.2	—	—	16.5*
30	—	18.6	19.6	17.3
40	18.2	—	—	17.7
		Pair-fed		
10	19.8	18.4	18.8	17.9

Enclosed figures are at maximal protein utilization.
* This value is apparently in error.

nitrogen × 6.25) of rats and mice are presented. The animals ate their food *ad libitum*, and varying amounts of different proteins were incorporated in the diets. Although there is considerable variation in carcass protein among the various groups, it will be seen that, in both species, the carcass protein is essentially the same at the level of

TABLE 8

PERCENT PROTEINS (*N x 6.25*) IN MOUSE CARCASSES (MINUS INTESTINAL CONTENTS) AFTER RECEIVING EXPERIMENTAL DIETS FOR 20 DAYS

Protein in diet (approximate)	Whole egg	Casein	Wheat gluten
%	%	%	%
4	17.4	16.6	—
6	17.5	17.0	17.2
8	17.6	18.0	17.0
11	17.9	17.1	17.1
15	19.1	17.9	17.6
18	—	19.1	17.8
23	21.4	18.5	17.7
35	18.1	18.4	18.4

☐ Enclosed figures are at maximal protein efficiencies

maximal protein utilization, regardless of the quality of the protein being ingested. From these data, it is concluded that total body weight provides a good index of body protein, if measurements are made at the level of protein intake providing maximal protein utilization.

B. MAINTENANCE

The above discussion has been restricted to certain factors that are involved in the growth measurements of protein quality. The remainder of this discussion will be devoted to factors that are commonly considered to be inadequately understood, in the measurement of the maintenance utilization of protein.

The Measurement of "Metabolic" Fecal Nitrogen

The usual procedure for measuring the metabolic fecal nitrogen is to determine the amount of nitrogen in the feces of animals receiving a protein-free diet. There is some question as to whether nitrogen excretion, during a protein-free period, measures so-called metabolic fecal nitrogen under conditions of protein feeding. In FIGURE 7, is shown the fecal nitrogen of mice, plotted as a function of the protein content of

the diet. The fecal excretion is expressed as mgms. nitrogen per 100 gms. of dry food consumed, in accord with Schneider's[9] observation that, within certain limits of food consumption, the metabolic fecal nitrogen is a function of the dry food intake. It will be seen that there is a

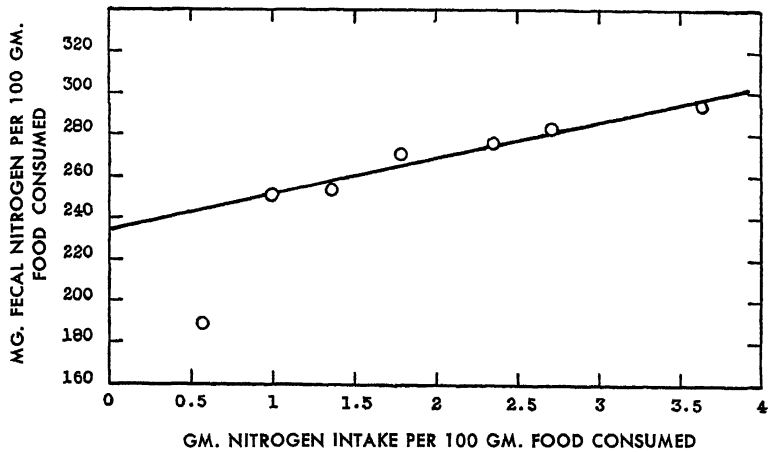


FIGURE 7 Fecal nitrogen excretion of mice receiving increasing amounts of protein (casein). (**Bosshardt, D. K., & H. H. Barnes.** J Nutrition **31:** 13 1946)

regular progression in fecal nitrogen, as the protein content of the diet increases. However, with the lowest protein diet, which, incidentally, was so low that it would not support growth, there is a marked deviation in fecal nitrogen from this regular progression In three different studies with mice, when the intake of protein was low enough to prohibit growth, this same observation has been made. An extrapolation of the fecal nitrogen figures, for the mice receiving protein in the diet to the point of zero protein intake, will give a considerably different value for the metabolic fecal nitrogen than will be obtained by the direct method of measuring fecal nitrogen, during the feeding of a protein-free diet. The influence of this difference in the metabolic fecal nitrogen on the calculation of the biological value of protein is obvious. In most measurements of the biological value of proteins, the period of protein-free feeding is relatively short, and, sometimes, a small amount of egg protein is incorporated in the diet. It may be that the error indicated in the mouse studies would not be a factor, during the short period of feeding a protein-free diet. However, these results are sufficiently suggestive to warrant further investigation of the validity of measuring metabolic fecal nitrogen by feeding a protein-free

diet. Furthermore, results of both rat and mouse studies have indicated that egg protein is not completely absorbed, and it may be that the ingestion of small amounts of this protein source, during the measurement of metabolic fecal nitrogen, is not desirable.

Measurement of Endogenous Urinary Nitrogen

One of the most frequent criticisms of the biological value method of measuring the nutritive quality of proteins, concerns the validity of the endogenous urinary nitrogen measurement. It is likely that much of this criticism is due to careless measurements of urinary nitrogen excretion of animals receiving a protein-free diet and, therefore, should be considered an error in technic, rather than validity. However, during the last year, there has been an accumulation of evidence, largely unpublished, from the laboratories of the Bureau of Biological Research, at Rutgers University, that may throw additional light on this problem. Allison and his associates[7] have found that the apparent biological value of proteins is markedly increased when the proteins are fed to dogs in a state of hypoproteinemia. Furthermore, an excellent protein source such as egg white, which, in a normal animal, has an apparent biological value of approximately 100, may far exceed this, when fed to the depleted animal. Some of these general observations have been recorded by others[10, 11] but, for the purpose of this discussion, specific reference will be made only to the work of the Rutgers group. An apparent biological value in excess of 100 means that, upon feeding a protein, the nitrogen excretion in the urine decreases. If the calculated biological value is exactly 100, there will be no change in urinary nitrogen, upon ingesting the protein. It is obvious that the biological value, as it is defined by Thomas, cannot exceed a value of 100, and, therefore, apparent retentions in excess of 100 must mean a change in the so-called endogenous urinary nitrogen. The reason for a change in endogenous urinary nitrogen, in a protein-depleted animal, remains obscure. However, it is necessary to recognize this fact, if an accurate interpretation of the biological value of protein is to be given. The observation that the depleted animal exhibits a decreased urinary nitrogen, when fed egg protein, does not mean necessarily that, in a normal nutritional state, endogenous urinary nitrogen will change, under conditions of protein feeding. It is assumed that, in the usual measurement of the biological value, the animal is not subjected to conditions that will cause the development of protein depletion and that, throughout the course of the determination, the endogenous urinary nitrogen remains constant. In small animals, such as the rat and mouse, it has

been found that very short periods of feeding protein-free diets will result in a considerable depletion of the body stores of protein.[12] This may mean that, during the time necessary for the attainment of urine nitrogen constancy, on a protein free diet (6–8 days), the small laboratory animal may develop a state of protein depletion that would be sufficient to alter endogenous urinary nitrogen excretion.

In many such studies, a small amount of whole egg protein is incorporated in the diet, rather than leaving it completely devoid of protein. The amount of egg protein is usually about sufficient for the maintenance of nitrogen equilibrium, and, as it is assumed that the biological value of this protein is essentially 100, the urinary nitrogen, during this regime, may represent endogenous excretion. In the small animal studies where egg protein diets have not been used, it is possible that one should review the data carefully, in order to determine whether or not sufficient protein depletion has taken place to result in an altered endogenous urinary nitrogen. In human protein nutrition studies, Murlin and his associates[13] have made a practice of feeding egg protein to their subjects, in place of the protein-free diet. The original reason for this replacement diet was to overcome the unpalatability of the protein-free regimen. In their early studies, milk protein was used for this purpose,[14] but, when it was recognized that egg protein had a higher nutritive value, the latter protein was used.

It is probable that a true understanding of endogenous urinary nitrogen excretion, under conditions of protein feeding, will not be realized, until tracer studies with labelled protein are carried out. However, it would appear likely that, under carefully controlled conditions, accomplished by either feeding small amounts of egg protein in small animal studies or sufficiently short periods of protein-free feeding in large animals, measurements of endogenous urinary nitrogen can be made, that will enable one to calculate the true biological value of proteins.

The Measurement of the Combined Utilization of Protein for Growth and Maintenance

Mitchell[15] extended the biological value measurement to growing animals, and it is with this type of test animal that most determinations are now carried out. Little attention has been given to the relative utilization of the dietary protein, for purposes of maintaining the nitrogenous integrity of the body tissues and in providing for the gain of body protein. It is recognized that the protein requirements for these two processes are not the same. Therefore, it is important to have a clear understanding of the relative proportions of dietary protein that

are diverted into these two types of utilization. By using the method originally employed by Osborne and Mendel,[16, 17] it has been possible to measure the amount of protein that is utilized for the maintenance of the growing rat. This is done by feeding amounts of protein that will maintain nitrogen constancy of the young rats, without permitting

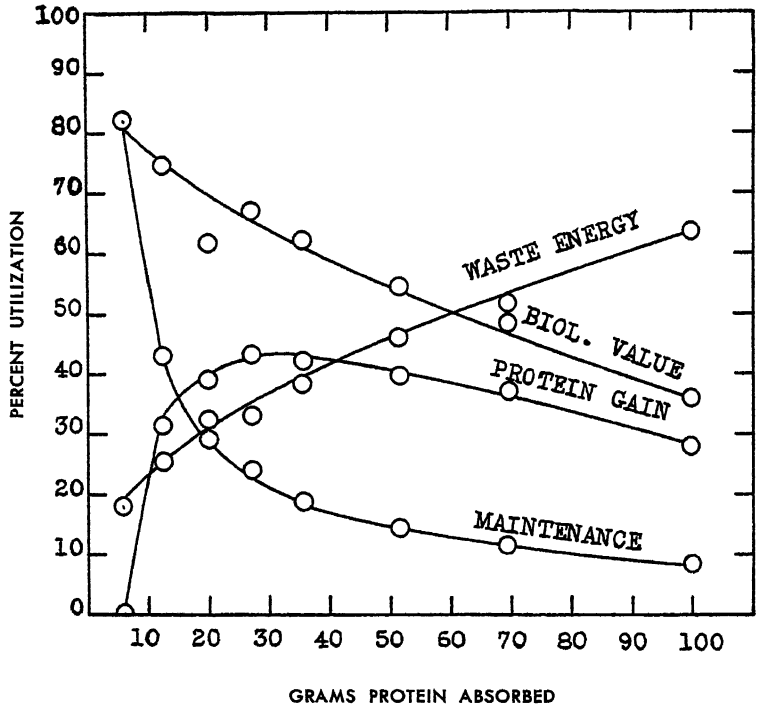


FIGURE 8. The comparative utilization of absorbed protein for maintenance and growth. The protein source is a well heat-treated soyflour and was fed *ad libitum* to groups of rats over a 42 day period. Waste energy is the absorbed protein that was not utilized for growth or maintenance and is so called because this caloric utilization represents a dispensable function of proteins.

growth to take place. The adequacy of this maintenance measurement has been checked by measuring the maintenance requirements of adult rats, by the method of Melnick and Cowgill.[18] When expressed on the basis of body surface area, it was found that the maintenance requirements, as measured by these two methods, checked. This indicates that the maintenance requirement is a function of body size and, in this sense, confirms the results of Smuts,[19] Olsen and Palmer,[20] and others. In FIGURE 8, the relative utilization of absorbed protein for maintenance and growth is shown. The protein source used for this

illustration is a high quality soyflour, but this same type of calculation has been applied, also, to three other protein sources. It will be seen that the percentage of absorbed protein that is utilized for carcass gain passes through a maximum, as the amount of absorbed protein increases. The percentage of absorbed protein that is used for maintenance reaches its highest point at that level of absorbed protein that maintains the animal without any growth taking place. If more protein is ingested, the percentage of absorbed protein utilized for maintenance decreases. The sum of the growth and maintenance factors is the biological value. As a result of the changes in growth and maintenance utilization, the biological value decreases, as more protein is absorbed. This decrease in biological value, with increases in food consumption, has been observed by Mitchell.[21] It may be pointed out, here, that the maintenance measurement that is carried out, by the method of Allison and Anderson,[22] is in the region of negative nitrogen balance that is represented by protein levels less than that level which provides for maintenance alone. It is assumed that the percentage of absorbed protein that is used for maintenance in this region of protein absorption will remain constant. For purposes of this discussion, the pertinent fact is that, in the region of positive nitrogen balance, at different levels of protein intake, the relative amounts of absorbed protein used for growth and maintenance are not constant.

A study of three other proteins having different nutritive qualities has shown that, at a given percentage protein in the diet, the relative utilization of absorbed protein, for growth and for maintenance, is different for each protein. In TABLE 9, is presented a comparison of the

TABLE 9

RELATIVE PARTICIPATION OF GROWTH AND MAINTENANCE IN THE ESTABLISHMENT OF THE BIOLOGICAL VALUE WHEN APPROXIMATELY 10 PER CENT PROTEIN IS INCLUDED IN THE DIET

Protein source	Protein apparently absorbed per rat/ 42 days	Biological value	Relative participation of growth	Relative participation of maintenance
	gm.	%	%	%
Whole egg	36	99	77	23
Soyflour No. 1	27	67	65	35
Soyflour No. 2	20	64	59	41
Wheat gluten	20	35	26	74

four protein sources that were studied. The wide range of participation of the absorbed protein, in growth and maintenance, is clearly evident. Since it is well recognized that the protein requirements are considerably different for maintenance and growth, the interpretation of the biological value, in the growing animal, becomes very confusing. The conclusion that is drawn from this study is that growth and maintenance determinations should be made separately, if a clear interpretation of the biological value is to be obtained.

CONCLUSIONS

Growth

1. Growth rates should not be used alone, as an index of the nutritive quality of proteins.

2. In the measurement of the growth-promoting qualities of protein, the level of protein in the diet has a considerable effect upon the results that are obtained. The maximal ratio of weight gain to protein ingested should be obtained. At this maximal ratio, the amount of protein absorbed is essentially the same, regardless of the nutritive quality of the protein. The common practice of measuring gain in body weight, as a function of protein consumption, with diets fed *ad libitum* and containing an arbitrarily selected amount of protein, may be useful in a preliminary rating of proteins, but should not be given wide application, without a full understanding of its limitations.

3. The protein efficiency values that are obtained by the method of paired-feeding do not coincide with the maximal protein efficiency ratios that may be obtained by *ad libitum* feeding. If the maximal ratio of weight gain to protein ingested is considered to be the most valid expression of the nutritive quality of proteins, for growth, the paired-feeding method should not be employed.

4. Evidence has been presented that the influence of caloric intake upon the growth utilization of proteins may be greater than has been suspected, heretofore. It is probable that this is particularly important when the protein content of the diet is low enough to result in a marked restriction of food intake. There is some evidence that, even in the range of maximal protein efficiency ratios, caloric consumption may not be optimal.

5. Under conditions of protein feeding where the maximal protein efficiency ratio is obtained, total body weight provides a good index of total body protein.

Maintenance

1. The nitrogen excretion, in the feces of an animal that is maintained on a very low protein diet, over a long period of time, is not indicative of "metabolic" fecal nitrogen, under conditions of more adequate protein feeding. In the small animal maintained for short periods of time on a protein-free diet, fecal nitrogen may not represent the "metabolic" fecal nitrogen of a protein-feeding period.

2. Urinary nitrogen excretion, during the period of feeding a protein-free diet to a protein-depleted animal, does not represent "endogenous" urinary nitrogen, for conditions of protein feeding. Further emphasis is given to the feeding of a small amount of egg protein, rather than a protein-free diet, for the measurement of "endogenous" urinary nitrogen in small animals.

3. The relative utilization of dietary protein, for growth and maintenance, is dependent upon the level of protein ingested and the nutritive quality of the protein. Since the protein requirements for growth and maintenance are different, it is concluded that these two factors should be measured independently and not in combination.

BIBLIOGRAPHY

1. **Osborne, T. B., L. B. Mendel, & E. L. Ferry**
1919. J. Biol. Chem. **37**: 223.
2. **Osborne, T. B., & L. B. Mendel**
1916. J. Biol. Chem. **26**: 1.
3. **Mitchell, H. H., & J. R. Beadles**
1930. J. Nutrition **2**: 225.
4. **Lusk, G.**
1928. The Science of Nutrition: 352. W. B. Sanders. Philadelphia.
5. **Allison, J. B., J. A. Anderson, & R. D. Seeley**
1946. Ann. N. Y. Acad. Sci. **47** (3): 273
6. **Larson, P. S., & I. L. Chaikoff**
1937. J. Nutrition **13**: 287.
7. **Mitchell, H. H.**
1944. Ind. & Eng. Chem. Anal. Ed. **16**: 696.
8. **Beadles, J. R., J. H. Quisenberry, F. I. Nakamura, & H. H. Mitchell**
1933. J. Agr. Research **47**: 947.
9. **Schneider, B. H.**
1935. J. Biol. Chem. **109**: 249.
10. **Willman, W., P. P. Swanson, G. F. Stewart, G. T. Stevenson, & M. Brush**
1945. Fed. Proc. **4**: 164.
11. **Peters, J. P.**
1944. Fed. Proc. **3**: 197.
12. **Bosshardt, D. K.**
Unpublished data.
13. **Murlin, J. R., E. B. Nasset, & M. E. Marsh**
1938. J. Nutrition **16**: 249.

14. **Murlin, J. R., & H. A. Mattill**
1938. J. Nutrition **16**: 15.
15. **Mitchell, H. H.**
1923. J. Biol. Chem. **58**: 873.
16. **Osborne, T. B., & L. B. Mendel**
1915. J. Biol. Chem. **20**: 351.
17. **Osborne, T. B., & L. B. Mendel**
1915. J. Biol. Chem. **22**: 241.
18. **Melnick, D., & G. R. Cowgill**
1937. J. Nutrition **13**: 401.
19. **Smuts, D. B.**
1935. J. Nutrition **9**: 403.
20. **Olsen, F. C., & L. S. Palmer**
1940. J. Agr. Res. **60**: 331.
21. **Mitchell, H. H.**
1924. J. Biol. Chem. **58**: 905.
22. **Allison, J. B., & J. A. Anderson**
1945. J. Nutrition **29**: 413.

DISCUSSION OF THE PAPER

Dr. Walter C. Russell (*Department of Agricultural Biochemistry, Rutgers University, New Brunswick, N. J.*):

As Dr. Barnes has indicated, basically two methods are in use in attacking the problem of the nutritive value of protein. One involves the measurement of the increase of cell mass, with or without relation to the quantity of protein consumed and the composition of the tissue produced; the other, the quantity of nitrogen retained by the cell mass from that absorbed from the digestive tract. The results obtained with these methods have permitted a classification of proteins, according to nutritive or biological values, so that proteins may be said to be of excellent, medium, or poor quality. Aside from serving, to some extent, as a guide in compounding rations for farm animals and diets for the human, the results of these methods of the evaluation have been of little value in answering the question as to what combinations of proteins should be used to obtain optimum response, in farm animals and in humans. For example, the compounding of rations for the efficient production of swine is still on an empirical basis. Despite the number of experiments reported, and still being conducted, there is not, as yet, an understanding as to why certain food combinations are effective and others are not. Furthermore, the experimental methods used, to date, have not furnished a basis for predicting what protein combinations should be used, to give a complete pattern of amino acids for all requirements of the animal.

I should like to report briefly some of the attempts we are making to use the white rat for a bioassay procedure In a study of the nutritive value of legumes cooked for human consumption, the method of paired-feeding for equal consumption was used, with skimmed milk powder as the reference protein. During a 40-day trial, the animals which received pea bean protein gained an average of 0.5 gm., and consumed an average of 5.5 gm. of food, per day. Those receiving the skimmed milk protein consumed the same quantity of food and gained about 1 gm. per day. On an *ad libitum* basis, the animals on the skimmed milk diet could gain 2.7 gm., and consume upward of 10 gm. of food, daily. It was apparent, as Dr. Barnes has pointed out, that the animals on the better protein were being severely penalized. These results told us only that one protein was better than another, but gave us no information as to how the poorer protein differed from the better.

Although Dr. Rose's studies at Illinois (Science, 86: 298, 1937) indicate that all of the essential amino acids must be present, if one is to obtain a growth response, the feeding of one essential amino acid to each of the rats which had been on the legume diet for 40 days was tried, in an attempt to find what essential amino

acids were lacking. Only in the case of methionine was there a response. This was quite dramatic, in that the animal gained 6 gm., in the first 24 hours, and an average of 2.4 gm. per day during a 12-day test period. The addition of all essential amino acids, except methionine, gave no response. When methionine was removed from the diet, there was a sharp drop in weight, there being a loss of 13 gm. in 7 days, or almost 2 gm. per day. This observation of the response of the legume to the methionine supplementation is not new, since it has already been reported by Woods, Beeson, and Bohn (J. Nutr. 26: 327, 1943) and others. We were impressed, however, by the rapidity of the response to methionine feeding, and the abrupt fall in weight when the amino acid was removed from the diet, indicating that there was no storage of this nutrient factor, as in the case of the vitamins.

The same study was conducted with yellow corn meal as the source of protein. In this instance, there was no marked stimulation of growth, but slight increases were noted for lysine, phenylalanine, tryptophane, and valine. In the case of gelatin, the use of the single essential amino acids failed to promote growth. When a group of five of them, picked at random, was fed, there was no response and, when the remaining five were fed, they failed to stimulate growth. On the other hand, when all of the ten amino acids were fed with gelatin, the latter at the 8% protein level, there was a marked increase in growth rate.

On the basis of these results, one could draw the tentative conclusion that the protein of pea beans furnished a considerable percentage of each of the essential amino acids, except methionine, whereas the protein of corn and gelatin did not provide a pattern of essential amino acids, which, when supplemented with single amino acids, would stimulate growth. Thus, two proteins could show, by growth or balance methods, a low nutritive value. The deficiency in one may be due to the lack of single amino acid, whereas, in the other, it might result from inadequate quantities of several. In compounding a diet, the first protein might be readily supplemented; the other, with difficulty.

It is well known that the protein of the soy bean promotes good growth in the white rat. In our laboratory, as previously demonstrated by Dr. Helen S. Mitchell (Western Hospital and Nurses' Rev. 11: 26, 1928), the protein of the chick pea, also called garbanza, ceci, and gram, promotes growth almost equal to that of the soy bean. Both of these legumes have a protein of higher nutritive quality than any of the common legumes so far studied.

Since, with none of the common legumes fed as a sole source of protein, is there more than slight growth stimulation, it seemed worth while to give attention to the question as to whether certain varieties of the common legumes have better growth-promoting properties than others, and whether any of them equals the nutritive value of the proteins of soy beans and chick peas. Rather than enter upon an extensive study, on the basis of paired feeding methods, we used the following technique. The rapid response to the feeding of methionine, and the immediate loss in weight when this amino acid was removed from the diet, led us to try a short experimental period, as compared with the longer 4 to 8 week periods, which have been used so frequently. Cooked dried legume seeds were fed at the 10% protein level for a period of 10 days, following which the diet was supplemented by the incorporation of 0.1% methionine, for a 10-day period; after this, the supplementation was increased to 0.6%, for another 10 days. Following this period, the methionine was removed, and the effect noted when the basal ration only was fed for a 10-day period.

Seventeen varieties of peas, lima beans, and snap beans were studied. During the first 10 days of the feeding of the basal diet, the King of the Garden variety of lima beans showed a gain of 0.7 gm. per gram of protein consumed. When this variety was supplemented with 0.1% methionine, the gain per gram of protein consumed was 3 gm., and when the supplement was raised to 0.6%, the gain was 3.1 grams. During the last ten days, when the basal diet was fed, the animals actually lost weight. For this variety, it would seem that methionine is the principal, if not the only amino acid, lacking. This variety, the best of any tested, does not approach the proteins of the soy bean and chick pea, in growth-promotion. For all of the other varieties, there were either very slight gains or loss, during the basal period, and lesser gains, in each of the periods of supplementation, than those noted for the King of the Garden variety.

Unless methionine acts in some special way, it seems to be the principal deficiency of these legumes. On the other hand, the amino acid patterns of most of them are certainly not as complete as that of the King of the Garden variety, because one variety of peas, Thomas Laxton, gave a response of 1.7 gm., when 0.1% methionine was fed, and only 1.9 gm., when 0.6% methionine was the supplement. If each of the seventeen varieties had been tested by the paired feeding method, the conclusion would have been reached that the protein is of poor quality. Except for the numerical expression of nutritive value, on a relative basis, the use of the 10-day test period leads to the same conclusion.

These, then, are some attempts in the direction of shortening the period of assay of proteins and of endeavoring to use the rat for the purpose of determining amino acid deficiencies.

It seems to me that the fundamental question in this field of investigation is what information should we have, in order to predict the protein or combination of proteins necessary to meet the requirements of the animal organism, for all functions and at all stages of life. The ideal answer would be the complete and accurate picture of the essential amino acid composition of each protein; that is, the quantitative essential amino acid pattern. This would be the answer, if we could be sure that there are no other nutrient factors, as yet unknown, associated or not associated with protein, which are necessary for the optimum utilization of the amino acids absorbed from the digestive tract.

The essential amino acid contents of a number of proteins have been determined by chemical and microbiological methods, and data of this type are increasing in volume, but agreement between the results obtained by different methods is not always satisfactory. It would seem that the next step would be to use a small experimental animal, the rat or mouse, in an attempt to confirm, biologically, the amino acid patterns obtained by chemical and microbiological methods. Experiments dealing with this problem are now under way in our laboratories.

THE ELECTROPHORETIC STUDIES ON THE EFFECT OF PROTEIN DEPLETION ON PLASMA PROTEINS AND THE REGENERATION OF PLASMA PROTEINS AFTER ORAL ADMINISTRATION OF HYDROLYSATES PREPARED FROM CASEIN AND LACTALBUMIN*

By Bacon F. Chow

Division of Protein Chemistry, The Squibb Institute for Medical Research, New Brunswick, New Jersey

Electrophoretic analysis is a very valuable tool for biochemical research on proteins and has been used with success, not only as a guide to the purification and characterization of proteins,[1-4] but also for the detection of changes in plasma proteins, due to immunization or other physiological reactions. The earlier work of Tiselius[5] demonstrated that the production of antibodies, in response to immunization of rabbits, was accompanied by an increase of the gamma-globulin fraction of the electrophoretic patterns of their plasma. The results of other investigators[6-10] disclosed that antibodies produced by different species of animals, upon immunization, migrate either with the gamma-globulin fraction of serum, or as a new component, with a mobility intermediate between beta- and gamma-globulins.

Moore and his associates[11, 12] reported that the sera of intact, normal rats contain little or no protein component, usually designated as alpha-globulin, when electrophoretically analyzed in a saline-phosphate buffer at pH = 7.4. However, this component was shown to be present in the serum of every hypothyroid rat examined, no matter whether the hypothyroidism was due to thyroidectomy or thiouracil feeding. On the other hand, Lewis and McCullagh[13] reported that their hypothyroid patients had a low alpha-globulin, but an increased beta-globulin. In the cases of hyperthyroidism, the plasma albumin was below normal level, but, in the majority of cases, the alpha-globulin was markedly increased. After correction of hyperthyroidism by thyroidectomy, or of hypothyroidism by thyroid therapy, the plasma protein pictures approximated the normal.

Numerous investigators[14-16] tried to ascertain whether there was any

* The author wishes to express his gratitude to Shirley M. Quaid for technical assistance.

correlation between the electrophoretic patterns of plasma of patients and different diseases. They found that diseases such as pneumonia, malaria, and streptococcus infections caused a rise in the alpha-globulin fraction. Dole and Emerson[17] studied electrophoretic changes in the plasma protein patterns of patients with relapsing fever. They found that these patients had normal total protein concentrations, but a depression in albumin and increases in the globulins, notably fibrinogen and gamma-globulin.

Dole, Watson, and Rothbard[18] made similar studies in patients with scarlet fever. From their results, they concluded that there was an early depression of albumin, with rises in the $alpha_1$- and $alpha_2$-globulins and a delayed rise in gamma-globulins, in the pattern of plasma of both rheumatic and non-rheumatic subjects. It is generally agreed, therefore, that electrophoretic analysis is of no diagnostic value, because of the absence of any specific change in the pattern of patients with different types of diseases, and, in some cases (malaria, for instance), the patterns tend to return to normal, despite persistent infection. However, it is probable that determination of the total protein concentration and of the true albumin:globulin ratio, by electrophoretic analysis, would be a useful adjunct in evaluating the severity of diseases.

It is our belief that, in some cases, the alteration of plasma components is related, in part, to protein depletion, as well as to other metabolic shifts attributed to the disease.

THE EFFECT OF DEPLETION ON PLASMA PROTEINS

The effect of protein depletion on plasma protein components of dogs was investigated by Zeldis and Alling.[19] They found that, during the first twenty-four hours following acute plasmapheresis, appreciable quantities of all components of plasma proteins entered the circulating blood stream, even during fasting, to replace the proteins lost by the body. These results showed that the body has a store for all these proteins for replenishment. Alpha- and beta-globulins, according to their report, continued relatively elevated, during subsequent days. Initial albumin levels were regained more slowly than those of total globulin.

In a more recent paper, Zeldis[20] and his associates reported their electrophoretic studies on dogs, following chronic depletion of circulating plasma proteins. Their results showed that many weeks of low-protein feeding brought about the alteration of albumin and

globulin concentrations. The albumin concentration was decreased, whereas the alpha-globulin was increased. No significant changes occurred in the concentration of beta- or gamma-globulins or fibrinogen. Unfortunately, they did not determine the plasma volume in all the cases studied. Hence, the amount of circulating proteins can only be estimated.

We[21] have been interested in following the change of composition of various plasma proteins of dogs, during chronic depletion by both plasmapheresis and protein-free feeding. For this purpose, six dogs were depleted of protein, by feeding a protein-free diet, for 6 to 10 days, after which plasmapheresis was performed, for 2 to 5 days. During these days, one quarter of the blood volumes of the dogs was removed daily. The cells were returned as a suspension in a normal saline solution. This period of plasmapheresis was followed by protein-free feeding, until the plasma protein concentration was relatively constant, at 4.0 to 4.5 grams per 100 ml. of plasma.

The protein nitrogen, as well as the non-protein nitrogen, in plasma were determined by the micro Kjeldahl method. Plasma volume was determined by the method of Gregersen and Stewart.[22] For electrophoretic analyses, 10 ml. of plasma were diluted with an equal volume of 0.10 N sodium diethylbarbiturate buffer at pH $= 8.4 \pm 0.1$. The protein solution was then dialyzed in the ice-box, for at least 24 hours, against two liters of the same buffer. Electrophoretic analyses were carried out according to the technics of Longsworth.[23] His directions and methods of resolutions were also followed for the determination of relative percentage of various components. The ascending and descending patterns of the plasma of five dogs are shown in FIGURES 1a and 1b. The electrophoretic patterns in FIGURE 1a represent those plasma samples taken before depletion, and those in FIGURE 1b are the corresponding samples taken after depletion. These electrophoretic patterns of plasma of dogs usually showed a greater number of components than that of other species of animals, such as human, bovine, etc. There was no difficulty in separating fractions corresponding to albumin, alpha-globulins, and gamma-globulin. Their electrical mobilities show only slight variation, in all experiments. However, the separation of the other fractions was less distinct and, therefore, was more subject to error in resolution. A comparison of the area in FIGURE 1 showed distinctly that the area under the albumin peak decreased sharply, whereas the alpha-globulin peaks increased noticeably.

The percentages, as well as the total circulating alpha-globulins and

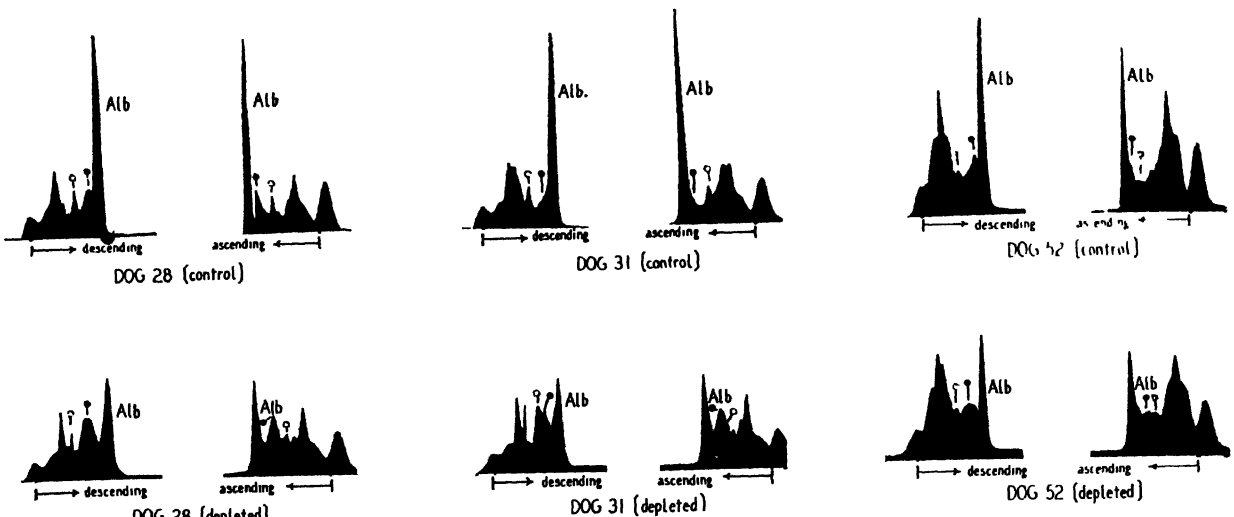


FIGURE 1a The descending and ascending patterns of plasma of three dogs before and after depletion (• = a_1 , ▫ = a_2)

albumins, are presented in TABLE 1. It was found (TABLE 1) that such depletion in dogs brought about hypoproteinemia, by the reduction of plasma volume (with the exception of Dog 54), as well as a lowering of plasma protein concentration. In terms of percentages, albumin was reduced, alpha-globulins were increased, and the gamma-globulin remained essentially unchanged. In terms of grams of total circulating proteins, albumin decreased markedly; the gamma-globulin tended to decrease slightly; whereas, the alpha-globulins showed very little change or even slight increase.

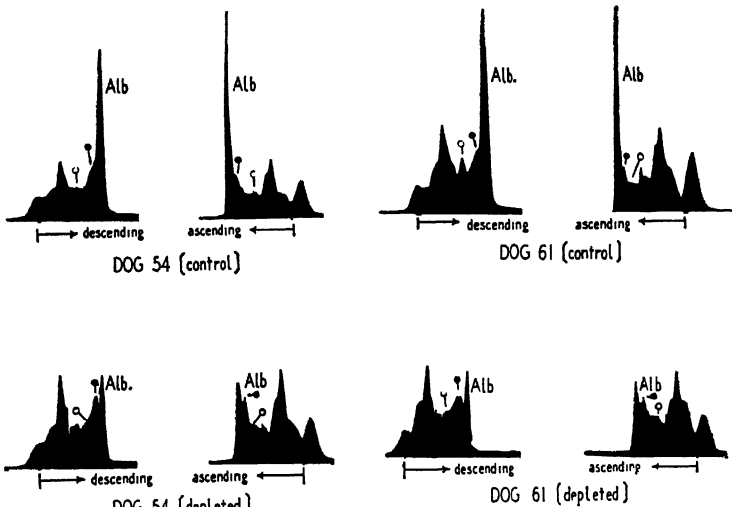


FIGURE 1b The ascending and descending patterns of plasma of two more dogs before and after depletion. ($\uparrow = \alpha_1$; $\downarrow = \alpha_2$.)

In order to find out the effect of protein depletion on other globulins, we have tabulated, in TABLE 2, the total circulating proteins, albumin, alpha-globulins, and other globulins, expressed in terms of grams per kilogram body weight. The term, "other globulins," is used to designate the protein components obtained by subtracting from the total protein both albumin and alpha-globulins. It can be seen that the net result of protein depletion on the other globulins is a marked decrease, in three out of six dogs; a slight increase, in one dog; and no change, in the remaining two. Further work is necessary to elucidate the difference in the behavior of these individual animals. The maintenance of circulating alpha-globulins, even in the highly depleted state, demonstrates the ability of the body to replace these protein components, which, presumably, play an important role in the normal functioning of the body.

TABLE 1

COMPOSITION OF PLASMA PROTEIN OF DOGS BEFORE AND AFTER PLASMAPHERESIS AND PROTEIN-FREE FEEDING

(Chow, Allison, Cole, & Seeley. Proc. Soc. Exp. Biol. & Med. 60: 14. 1945.)

Dog No.	Weight kg.	Plasma proteins gms./100 ml.	Plasma volume ml.	Plasma protein fractions, per cent				Total circulating plasma protein, gms.			Albumin/globulin		Treatment
				Albumin	Globulins			Albumin	Globulins		Howe*	Electrophoretic	
					α_1	α_2	γ		$\alpha_{1\&2}$	γ			
28	8.9	5.41	564	44	13	12	8	13.4	7.6	2.4	2.56	0.79	Control
	8.8	4.20	460	37	32	—	8	7.1	6.2	1.5	0.86	0.54	Depleted
31	5.8	5.96	431	42	6	9	9	10.8	3.9	2.3	1.32	0.72	Control
	5.9	3.84	352	27	32	5	9	3.7	5.0	1.2	1.21	0.37	Depleted
33	7.9	5.79	443	40	7	8	5	10.2	3.8	1.3	1.60	0.66	Control
	7.3	4.32	415	34	15	10	9	6.1	4.5	1.6	1.30	0.52	Depleted
52	9.5	6.36	442	29	13	8	12	8.2	5.9	3.4	0.77	0.41	Control
	9.0	5.15	388	24	19	9	14	4.8	5.6	2.8	0.60	0.32	Depleted
54	12.5	5.35	519	45	16	—	9	12.5	4.4	2.5	1.05	0.82	Control
	11.1	4.10	559	21	18	11	11	4.8	6.7	2.5	0.58	0.26	Depleted
61	12.3	5.67	566	43	9	11	8	13.8	6.4	2.6	1.26	0.75	Control
	11.0	4.40	498	16	20	16	10	3.5	7.9	2.2	0.56	0.19	Depleted

* Howe's fractionation method.

TABLE 2

CHANGES OF TOTAL CIRCULATING ALBUMIN AND GLOBULINS DUE TO PROTEIN DEPLETION

TCP = Total circulating proteins in gm kgm. body weight; TCA = Total circulating albumin in gm/kgm. body weight, TCαG = Total circulating alpha-globulins in gm/kgm. body weight; TCOG = Total circulating globulins, excluding α-globulins, in gm/kgm. body weight.

Dog. No.	Weight, kgm.	TCP	TCA	TCαG	TCOG
28	8.9	3.43	1.51	0.85	1.07
	8.8	2.73	0.81	0.83	1.07
31	5.8	4.44	1.86	0.66	1.92
	5.9	2.29	0.63	0.85	0.81
33	7.9	3.24	1.29	0.48	1.47
	7.3	2.46	0.84	0.62	1.00
52	9.5	2.96	0.86	0.62	1.48
	9.0	2.22	0.53	0.62	1.07
54	12.5	2.22	1.00	0.35	0.87
	11.1	2.06	0.43	0.60	1.03
61	12.3	2.61	1.12	0.52	0.97
	11.0	1.99	0.32	0.72	0.95

REGENERATION OF PLASMA PROTEINS

A. After the Administration of Casein Hydrolysate*

To bring about the regeneration of plasma proteins, we† have fed three dogs, which were previously so depleted by plasmapheresis and protein-free feeding, according to the method just described, that their plasma protein concentrations dropped from an initial value of about 6.5 grams per cent to about 4 grams per cent. Samples of their blood were taken, before and after depletion, as well as after each period of feeding of about 5 days, for the determination of plasma protein concentration and plasma volume, and for electrophoretic analysis.

The results of both descending and ascending patterns of plasma of Dog 62 are shown in FIGURE 2. These samples of plasma were drawn from the dog before and after depletion, as well as after different periods of feeding of casein hydrolysate. It can be noted that there was a decrease of albumin peak and an increase of alpha-globulin peaks, as a result of protein depletion (compare B with A). Five days after,

* Casein hydrolysate was prepared in the Squibb Institute for Medical Research, by digestion of casein with trypsin.

† **Chow, B. F., J. B. Allison, W. H. Cole. & R. D. Seeley.** In press.

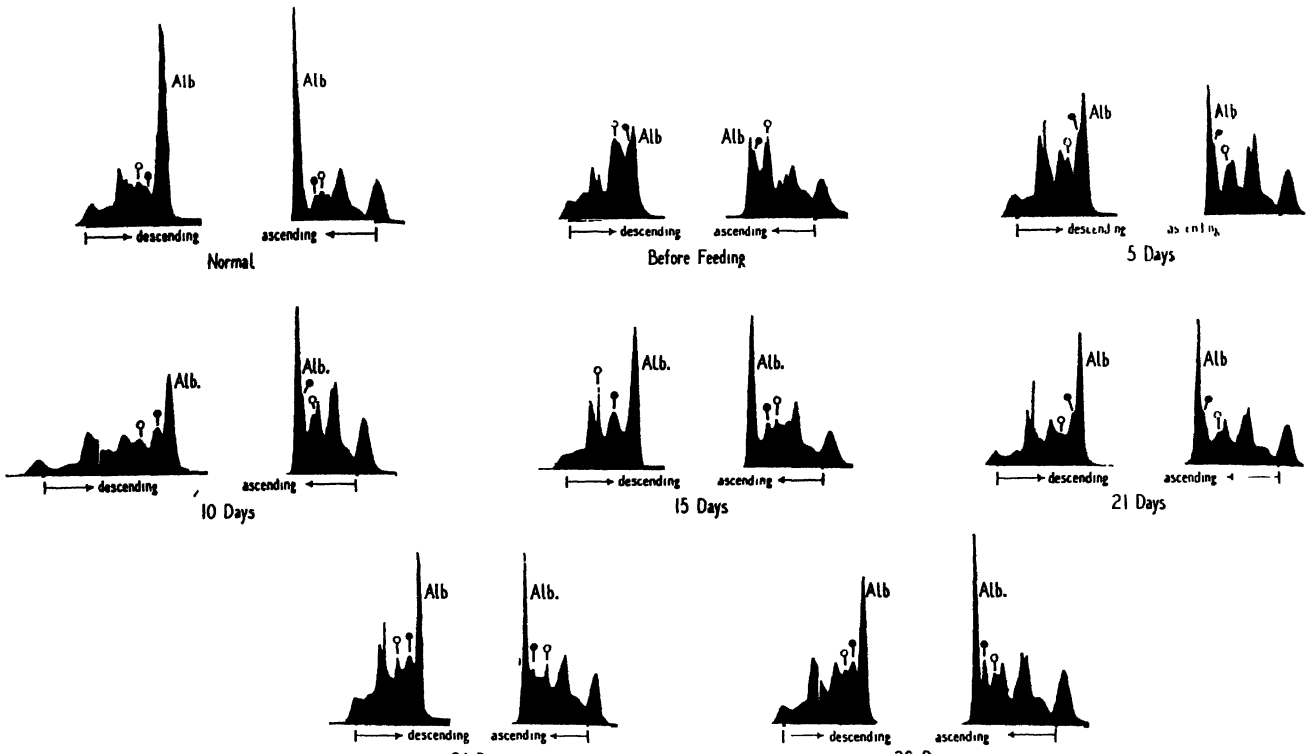


FIGURE 2. The descending and ascending patterns of plasma of dog 62, showing the effect of protein depletion and repletion after the oral administration of casein hydrolysate. ($\overset{\alpha_1}{\downarrow}$ = α_1; $\overset{\alpha_2}{\downarrow}$ = α_2)

oral feeding brought about a very marked increase in albumin content. After 36 days of feeding, the electrophoretic pattern was essentially normal.

The compositions of the plasma of all three dogs are collected in TABLE 3 and show that oral administration of casein hydrolysate stimulated the regeneration of albumin and globulin fractions of all three dogs. The increase in total circulating proteins was noticeable, after

TABLE 3

PLASMA PROTEIN REGENERATION OF DOGS, PREVIOUSLY DEPLETED BY PROTEIN-FREE FEEDING AND PLASMAPHERESIS, FOLLOWING ORAL ADMINISTRATION OF CASEIN HYDROLYSATE

TCP = Total circulating proteins in grams; TCA = Total circulating albumin in grams; TCG = Total circulating globulins in grams; — = Indicates no electrophoretic analysis.

Dog. No.	Date	TCP	TCA	TCG	Treatment
62	5/24	21.1	10.4	10.7	Normal
	6/25	13.7	5.0	8.7	Before feeding
	6/30	16.4	5.9	10.7	Five days feeding
	7/5	16.0	4.3	11.7	Ten days feeding
	7/10	18.2	6.6	11.6	Fifteen days feeding
	7/16	20.5	6.7	13.8	Twenty-one days feeding
	7/26	21.0	6.7	14.3	Thirty-one days feeding
	7/31	22.0	6.8	15.21	Thirty-six days feeding
Net Increase	7/31	8.3	1.8	6.5	Thirty-six days feeding
61	3/9	32.1	13.8	18.3	Normal
	4/13	22.0	3.5	18.5	Before feeding
	4/18	30.4	—	—	Five days feeding
	4/23	35.0	9.1	25.9	Ten days feeding
	4/28	34.0	11.3	22.7	Fifteen days feeding
	5/3	31.3	11.3	20.0	Twenty-one days feeding
	5/8	41.0	12.6	28.4	Thirty-one days feeding
	5/14	46.6	16.9	29.7	Thirty-six days feeding
Net Increase	5/14	24.6	13.4	11.2	Thirty-six days feeding
54	3/6	27.7	12.5	15.2	Normal
	4/13	22.9	4.8	18.1	Before feeding
	4/17	29.5	5.9	23.6	Five days feeding
	4/23	32.2	8.4	23.8	Ten days feeding
	4/28	31.0	8.1	22.9	Fifteen days feeding
	5/3	35.2	14.8	20.4	Twenty-one days feeding
	5/8	34.1	11.0	23.1	Thirty-one days feeding
	5/14	36.0	13.7	22.3	Thirty-six days feeding
Net Increase	5/14	13.1	8.9	4.2	Thirty-six days feeding

the first period of feeding, *i.e.*, five days, for all three dogs. In order to bring the total circulating proteins back to the "normal" value, it required about one month of feeding, for Dog 62, and about 5 to 10

days, for the other two dogs. The difference in the rate of recovery to "normal" may be due, in part, to differences in the degree of protein depletion and the variation of response among dogs.

It is of interest to note that oral administration of the casein-hydrolysate rapidly brought about the regeneration of both albumin and globulins, and, after continued feeding, the total circulating albumin reached the "normal" values in two dogs (Dogs 61 and 54), but below the "normal" value in Dog 62. Dog 62 showed another abnormality, in that its total circulating albumin was decreased on 7/5 sample. However, sample of its plasma, taken on or after 7/10, again showed definite and consistent increase in albumin. Continued feeding of this hydrolysate raised the total circulating globulins considerably above the "normal" values in all three dogs. The net increase in total circulating proteins (albumin as well as globulins), which is obtained by subtracting the total amounts in circulation, before feeding, from the corresponding amounts after thirty-six days of feeding, is given on the last line of TABLE 3. In all cases, there was regeneration both of albumin and globulins, although the ratio of regenerated albumin to regenerated globulins showed a wide variation among the three dogs (0.20, Dog 62; 1.17, Dog 61; 2.12, Dog 54).

B. After the Oral Administration of the Lactalbumin Hydrolysate*

Two of the protein-depleted dogs were fed the lactalbumin** hydrolysate orally. Samples of the plasma were obtained and analyzed, in the same manner as that described for the other three dogs receiving casein hydrolysate. The results of both descending and ascending patterns of plasma of Dog 63 are shown in FIGURE 3. The electrophoretic patterns of the second dog yielded essentially the same results. These samples were withdrawn from the dog before and after depletion, as well as after different periods of oral administration of lactalbumin hydrolysate. It can be seen, from FIGURE 3, that oral administration of this hydrolysate brought about an increase in albumin and globulins, after 5 to 10 days of feeding. However, unlike the plasma of dogs receiving casein hydrolysate, the plasma of this dog showed more marked increase in albumin, after 36 days of feeding (compare hydrolysate of FIGURE 2 with hydrolysate of FIGURE 3). The analytical results of the composition of the plasma samples of both dogs are collected in TABLE 4.

* The lactalbumin hydrolysate used was likewise prepared by the tryptic digestion of lactalbumin.

** **Chow, B. F., J. B. Allison, W. H. Cole, & R. D. Seeley.** In press.

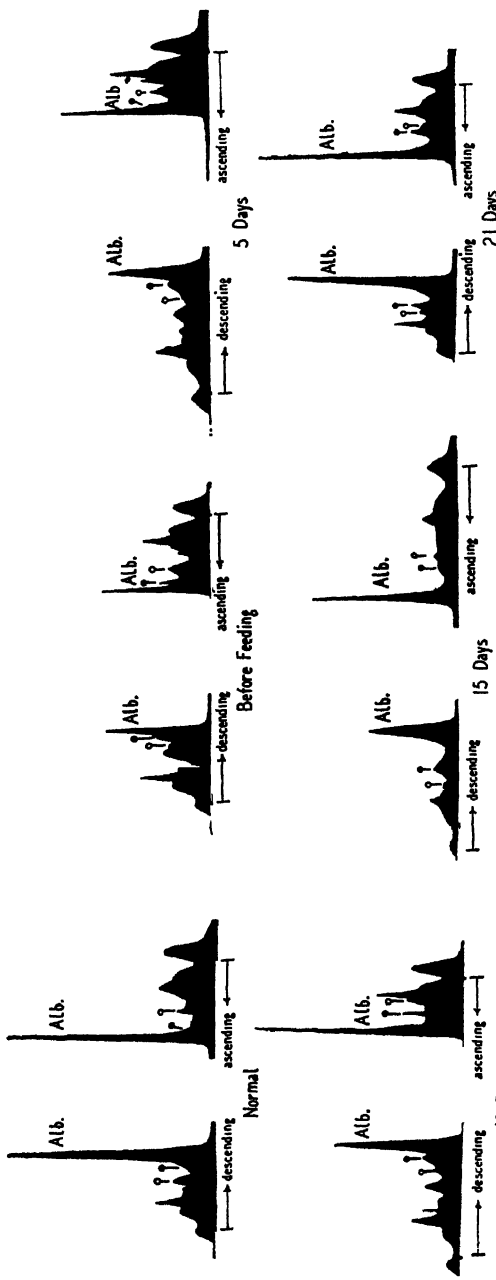


FIGURE 3. The descending and ascending patterns of plasma of Dog 63, showing the effect of protein depletion and repletion after the oral administration of lactalbumin hydrolysate. (? = α_1; ? = α_2.)

TABLE 1

PLASMA PROTEIN REGENERATION OF DOGS, PREVIOUSLY DEPLETED BY PROTEIN-FREE FEEDING AND PLASMAPHERESIS, FOLLOWING ORAL ADMINISTRATION OF LACTALBUMIN HYDROLYSATE

TCP = Total circulating proteins in grams; TCA = Total circulating albumin in grams; TCG = Total circulating globulins in grams.

Dog No.	Date	TCP	TCA	TCG	Treatment
63	5/25	29.7	13.6	16.1	Normal
	6/25	19.8	4.4	15.5	Before feeding
	6/30	24.3	5.8	18.5	Five days feeding
	7/5	30.9	10.8	20.1	Ten days feeding
	7/10	26.4	12.4	14.0	Fifteen days feeding
	7/16	27.6	12.7	14.9	Twenty-one days feeding
	7/26	28.8	15.9	12.9	Thirty-one days feeding
Net Increase	7/26	9.0	+11.5	−2.6	Thirty-one days feeding
60	5/24	26.8	11.5	15.3	Normal
	6/25	18.2	3.5	14.7	Before feeding
	6/30	23.0	4.1	18.9	Five days feeding
	7/5	25.3			Ten days feeding
	7/10	25.4	8.4	17.0	Fifteen days feeding
	7/16	28.0	11.8	16.2	Twenty-one days feeding
	7/26	30.1	14.4	15.7	Thirty-one days feeding
	7/31	27.9	14.2	13.7	Thirty-six days feeding
Net Increase	7/31	9.7	+10.7	−1.0	Thirty-six days feeding

Like the casein hydrolysate, the lactalbumin hydrolysate brought about an appreciable regeneration of total circulating proteins of these two depleted dogs, after five days of oral feeding. After about ten days of feeding, the regeneration of total circulating globulins appeared to have reached its peak, which was considerably higher than the normal value for both dogs. Further feeding of this hydrolysate caused a continual rise of albumin content, but a gradual decrease of globulin content. As a matter of fact, after about a month of feeding, the total circulating albumin was above "normal" value. The total circulating globulins receded to a level definitely below normal. This decrease** was due mainly to the lowering of alpha-globulin contents. Thus, the net increase of total circulating proteins (see the bottom line in TABLE 4) was due entirely to the increase of albumin.

A comparison of the results given in TABLE 3 and TABLE 4 shows that oral feeding of either casein hydrolysate or lactalbumin hydrolysate can bring about an increase in the total circulating albumin and globulins to their "normal" values. However, with prolonged feeding of lactalbumin hydrolysate, there is a trend of continued rise in albumin,

** To be published in another communication.

but a gradual decrease of globulins, so that the net increase in total circulating proteins appears to be entirely due to increase in albumin. On the other hand, with prolonged feeding of casein hydrolysate, the initial gain in both albumin and globulin fractions can be maintained, or even possibly increased, so that the net increase of total circulating proteins, after 36 days of feeding, is due to both albumin and globulins. Furthermore, this difference in plasma protein regeneration is reflected in the albumin:globulin ratio. Thus, this ratio for dogs receiving lactalbumin hydrolysate, for a month, was greater than one (1.23, Dog 63; 1.04, Dog 60), and it was less than one (0.28, Dog 62; 0.57, Dog 61; 0.61, Dog 54), for dogs receiving casein hydrolysate for 36 days.

Since a part, or all, of certain amino acids originally present in the proteins may be removed or destroyed, during the preparation of the hydrolysate, complete analyses of all essential amino acids, as well as some non-essential ones of the casein and lactalbumin hydrolysates, are being undertaken and will be reported elsewhere. Although the limited number of dogs does not permit us to draw definite conclusions, it appears certain that the oral administration of these two hydrolysates can regenerate different types of plasma proteins.

DEPLETION AND REGENERATION OF PLASMA PROTEINS IN MAN

A. The Relationship between Albumin and Alpha-globulins

From the above discussion, we note that protein depletion in dogs can bring about a decrease in percentage of albumin and an increase in percentage of alpha-globulins. It is of interest to ascertain whether such a relative change also takes place in man.

In collaboration with Dr. Co Tui, we have analyzed electrophoretically the plasma of 13 of his patients. Some of them were postoperative, while others suffered from tuberculosis and severe malnutrition. In all cases, the plasma protein concentration was also determined. The following points are worth noting: The patients, KA and ME (TABLE 5), were greatly depleted in the total protein concentrations (3.84 and 3.89 grams per cent, respectively). The albumin contents of their plasma were as low as 19 per cent and 24 per cent, but the alpha-globulins were much higher than normal, being 68 and 47 per cent, respectively. Similarly, the plasma of the patient, AL, with a total protein concentration of 5.78 grams per cent, contained only 24 per cent albumin, but as much as 38 per cent alpha-globulins. These re-

TABLE 5

ELECTROPHORETIC PATTERNS OF PLASMA* OF DR. CO TUI'S PATIENTS

Patient	Plasma proteins gms./100 ml.	Plasma volume ml.	Hematocrit	Tiselius analyses						Remarks
				Albumin	α_1	α_2	β	ϕ	γ	
LO	6.55	—	19	62	—	13	—	13	12	Appendectomy
BA	6.81	59.5	40	57	7	11	13	—	12	Gall bladder
HE	5.31	—	42	57	—	—	15	—	17	Bone fracture
SW	6.24	—	48	54	6	14	13	—	13	Hernia
WO	5.80	58.6	39	51	5	11	15	5	13	Ca-Stomach Malnutrition
CR	8.0	—	—	44	7	15	14	4	18	—
ZA	5.54	50.4	34	41	15	11	14	4	15	Malnutrition
HAL	8.0	—	—	37	10	13	16	6	18	—
BE	6.62	53.7	44	37	11	15	21	—	16	Hernia
HA	6.8	—	—	30	10	23	14	3	20	—
ME	3.89	—	36	24	22	25	—	16	13	Malnutrition
AL	5.78	—	32	24	14	24	24	—	14	Cancer
KA	3.84	—	42	19	4	64	4	1	7	Tuberculosis

* Composition of each component is expressed in per cent of the total plasma proteins.

sults suggest that patients with low plasma protein concentrations, that is, in the depleted state, also have less albumin, but more alpha-globulin, than normal. Hence, they appear to be in agreement with those results obtained in dogs. Furthermore, there is no apparent relationship between the decrease in per cent of albumin with the change of other plasma protein components, like beta- or gamma-globulins.

In collaboration with Dr. Tom Spies,* Miss M. Bieler, and Mr. R. Johnson, we have analyzed electrophoretically the plasma of 10 patients (TABLE 6) with diseases somewhat different from the former group. It is interesting to note that the patients, ED and RLP, had only 29 and 32 per cent plasma protein as albumin. Their alpha-

TABLE 6

ELECTROPHORETIC PATTERNS OF PLASMA* OF DR. TOM SPIES' PATIENTS

Patient	Serum proteins	Tiselius analyses, per cent						Disease
		Albumin	α_1	α_2	β	ϕ	γ	
ND	5.12	70	5	7	6	3	9	Control
EB	—	54	5	8	6	7	20	Addison's Disease
WC	6.97	50	4	6	9	3	28	Chronic monocytic leukemia
SJ	6.55	50	6	11	16	3	14	Diabetes mellitus
EMF	6.61	46	7	10	10	4	23	Rheumatic heart and epilepsy
AD	5.6	45	5	11	20	4	15	Diabetes mellitus
WB	—	45	9	15	9	9	13	Cerebral accident with arteriosclerotic heart
CD	5.55	43	6	11	19	4	17	Cardiac decompensation
RLP	5.89	32	9	18	12	4	25	Diabetes mellitus
ED	—	29	8	14	14	7	28	Tuberculosis

* Composition of each component is expressed in per cent of the total plasma proteins.

globulins were higher than normal (22 and 27 per cent, respectively). The results obtained from this group of patients again confirm the fact that a drop in albumin content is usually reflected by an increase in alpha-globulins.

From the results, it may be concluded that a drop in the albumin content is usually associated with a corresponding increase in the percentage of alpha-globulins. This relationship takes place for the plasma of patients, regardless of the cause of protein depletion (*i.e.*, low plasma protein concentration) and irrespective of the type of disease or history of the patient.

* In collaboration with the University of Cincinnati Studies in Nutrition at the Hillman Hospital, Birmingham, Alabama.

B. Plasma Protein Regeneration in Man

In order to study whether our casein hydrolysate will also regenerate plasma proteins, when given orally to man, we have fed it to four patients. The plasma protein concentration, as well as plasma volume, were determined before and after treatment. The albumin and globulin contents were determined electrophoretically. The results of such a series are collected in TABLE 7.

It can be seen that, in the first place, there was an increase in the total circulating plasma proteins, for all four patients. This increase is due both to an increase of plasma volume, as well as to plasma protein concentration, excepting the case of SI, who had an apparently high protein concentration of 7.13 grams per cent, but low plasma volume of 45.0 ml. per kilogram body weight, before the hydrolysate therapy. However, the decrease of plasma protein concentration, after 17 days of feeding, was more than compensated for by a greater increase of plasma volume, so that the net result was a definite gain in the total circulating proteins. Electrophoretic analyses showed that both albumin and globulin fractions were regenerated, following the protein hydrolysate therapy.

DISCUSSION

Changes in the composition of plasma proteins can take place as a result of immune reactions, infectious diseases, or diseases due to hormone deficiency, protein depletion, or physical injuries (burns, fracture). It appears, therefore, that, to study the effect of protein hydrolysate administration, it is of paramount importance to follow not only the qualitative, but also the quantitative, changes of the plasma proteins, in order to evaluate properly the efficacy of protein therapy.

Plasma protein regeneration studies have been made by numerous investigators. The most extensive investigations were made by Weech [24-26] and his associates. In their studies, these observers used intact proteins, such as serum, casein, and meat. Cox, Mueller, and their associates[27, 28] reported on the serum albumin regeneration, as effected by intravenously and orally administered hydrolysate, prepared by the pancreatic digestion of different proteins. For comparative measurement of serum protein regeneration, they used the technique developed by Weech and Goettsch.[25] Albumin regeneration was measured by determining the total protein concentration with the Kjeldahl method, and the percentage of serum albumin was determined by the method of Howe.[29] Thus, the change in serum albumin, during a week of sup-

TABLE 7

PLASMA PROTEIN REGENERATION OF PATIENTS FOLLOWING PROTEIN HYDROLYSATE ADMINISTRATION

TCP = Total circulating proteins; TCA = Total circulating albumin; TCG = Total circulating globulins.

Patient	Date	Gms. of hydrolysate/Kgm. body wt	Plasma proteins gms./100 ml.	Plasma volume ml.	TCP	TCA	TCG	Albumin/globulin	Disease
ST	4/17 4/27	4.4 SH	5.81 6.58	54.5 56.4	3.16 3 73	1.52 1.85	1.64 1.88	0.93 0.98	Peptic ulcer
*Increase of plasma protein**					0.57	0.33	0.24		
SI	4/25 5/12	4.4 SH	7.13 6.60	45.0 57.5	3 21 3.80	2 00 2.20	1.21 1.60	1.65 1.38	Gastric ulcer
*Increase of plasma protein**					0.59	0.20	0.39		
DA	6/29 7/24	6.5 SH	5.36 7.01	55.1 56.6	3.00 3.96	1.29 1.98	1 70 2.00	0.76 0.99	Malnutrition
*Increase of plasma protein**					0.96	0.69	0.30		
BE	11/19 11/26	4.4 SH	6.09 6.22	48.3 53.7	2.94 3.57	1.47 1.61	1.47 1.96	1.00 0.82	Hernia
*Increase of plasma protein**					0.63	0.14	0.49		

* In grams per kilogram body weight.

plementary feeding of the protein, was taken as the basis for comparison. Upon these results, it was concluded that enzymic hydrolysates of casein, lactalbumin, and beef serum proteins are equally effective in the regeneration of plasma albumin in hypoproteinemic dogs, whether given orally or intravenously.

Elman[9] and his co-workers made similar studies on regeneration of serum proteins with protein hydrolysates. They reported increases in serum protein concentration, in operative patients, with hypoproteinemia following intravenous administration of the hydrolysate. Regeneration of serum albumin also took place in protein-depleted dogs, after oral or intravenous administration of the hydrolysate. The albumin determination was also made, with the chemical fractionation method.

Any discussion of plasma protein metabolism would be incomplete without reference to the classic work of Professors Whipple and Madden. For their study, they prepared a "standard protein-depleted dog," by plasmapheresis and protein-free feeding. By limiting and controlling the diet nitrogen intake, they were able to bring these dogs to a steady state of hypoproteinemia and a constant level of plasma protein production. Administration of appropriate mixtures of essential amino acids brought about nitrogen retention, as well as protein regeneration. Omission of some one essential amino acid from such a mixture may make it totally inadequate for plasma protein regeneration.

In this communication, we have emphasized, not the change of protein concentration, but the change of circulating proteins, as a result of protein hydrolysate administration. The importance of the determination of plasma volume cannot be overemphasized, since regeneration of total protein may occur, even with a decrease of plasma protein concentration, because of larger increase of plasma volume. In addition, with the aid of electrophoretic analysis, one can not only determine more accurately the albumin:globulin ratio than with the chemical method, but one can follow the change of the various protein components, either during depletion or regeneration. In this way, it was possible to differentiate the two different patterns of plasma protein regeneration, after casein and lactalbumin hydrolysate therapy.

There appears to be a relationship between the drop in albumin and a corresponding rise in alpha-globulins. The rapidity with which alpha-globulins are replaced, after the removal of a large percentage of this fraction of proteins (by plasmapheresis and persistent increase in alpha-globulins, due to a variety of diseases), indicates their physiological importance, although their functions are not well understood.

SUMMARY AND CONCLUSIONS

It was found, with the aid of electrophoretic analyses, that protein depletion brings about a decrease in the circulating albumin and globulins (with the exception of alpha-globulins). Oral administration of protein hydrolysate to depleted animals brings about regeneration of plasma proteins. The casein hydrolysate will regenerate both albumin and globulins, even on prolonged feeding. The lactalbumin hydrolysate will also regenerate both albumin and globulins, at the start. However, prolonged feedings of this hydrolysate favor the regeneration of albumin, and result in a decrease of globulins.

Plasma of patients with different types of diseases, with total protein concentration ranging from 4 grams per cent to 8 grams per cent, were analyzed electrophoretically. The results indicate that protein depletion, due to malnutrition, tuberculosis, or cancer, is invariably accompanied by the lowering of the albumin and corresponding increases in alpha-globulins. Oral administration of our casein hydrolysate will regenerate both albumin and globulin fractions.

BIBLIOGRAPHY

1. **Tiselius, A.**
1937. Trans. Faraday Soc. **33**: 524.
2. **Pirie, N. W.**
1940. Biol. Rev. Cambridge Philos. Soc. **15**: 377.
3. **Shedlovsky, T.**
1943. Ann. N. Y. Acad. Sci. **43**: 253.
4. **Chow, B. F.**
1944. Advances in Protein Chemistry. **I**: 153. Academy Press. New York, N. Y.
5. **Tiselius, A.**
1937. Biochem. J. **31**: 1464.
6. **Pappenheimer, A. M., Jr., H. P. Lundgren, & J. W. Williams**
1940. J. Exp. Med. **71**: 247.
7. **van der Scheer, J., & R. W. G. Wyckoff**
1940. Proc. Soc. Exp. Biol. & Med. **43**: 427.
8. **van der Scheer, J., R. W. G. Wyckoff, & F. H. Clarke**
1941. J. Immunol. **40**: 173.
9. **Kekwick, R. A., & B. R. Record**
1941. Brit. Exp. Path. **22**: 29.
10. **Davis, B. D., E. H. Moore, E. A. Kabat, & A. Harris**
1945. J. Immunol. **50**: 1.
11. **Moore, D. H., L. Levin, & J. H. Leathem**
1944. J. Biol. Chem. **153**: 349.
12. **Moore, D. H., L. Levin, & G. K. Smelser**
1945. J. Biol. Chem. **157**: 723.
13. **Lewis, L. A., & E. P. McCullagh**
1944. Am. J. Med. Sci. **208**: 727.
14. **Longsworth, L. G., T. Shedlovsky, & D. A. MacInnes**
1939. J. Exp. Med. **70**: 399.

15. **Kekwick, R. A.**
1940. Biochem. J. **34**: 1248.
16. **Luetscher, J. A., Jr.**
1940. J. Clin. Invest. **19**: 313.
17. **Dole, V. P., & K. Emerson, Jr.**
1945. J. Clin. Invest. **24**: 644.
18. **Dole, V. P., R. F. Watson, & S. Rothbard**
1945. J. Clin. Invest. **24**: 648.
19. **Zeldis, L. J., & E. L. Alling**
1945. J. Exp. Med. **81**: 515.
20. **Zeldis, L. J., E. L. Alling, A. B. McCoord, & J. P. Kulk**
1945. J. Exp. Med. **82**: 157.
21. **Chow, B. F., J. B. Allison, W. H. Cole, & R. D. Seeley**
1945. Proc. Soc. Exp. Biol. & Med. **60**: 14–17.
22. **Gregersen, M. I., & J. D. Stewart**
1939. Am. J. Physiol. **125**: 142.
23. **Longsworth, L. G.**
1942. Chem. Rev. **30**: 323.
24. **Weech, A. A., E. Goettsch, & E. B. Reeves**
1935. J. Exp. Med. **61**: 299.
25. **Weech, A. A., & E. Goettsch**
1938. Bull. Johns Hopkins Hosp. **63**: 154, 181.
26. **Weech, A. A., & E. Goettsch**
1939. Bull. Johns Hopkins Hosp. **64**: 425.
27. **Cox, W. M., Jr., & A. Mueller**
1944. J. Clin. Invest. **23**: 875.
28. **Mueller, A. J., K. S. Kemmerer, W. M. Cox, Jr., & S. T. Barnes**
1940. J. Biol. Chem. **134**: 573.
29. **Howe, P. E.**
1921. J. Biol. Chem. **49**: 109.
30. **Elman, R.**
1940. Ann. Surg. **112**: 594.
31. **Madden, S. C., R. R. Woods, F. W. Shull, & G. H. Whipple**
1944. J. Exp. Med. **79**: 607.

BLOOD PROTEIN REGENERATION AND INTERRELATION

By G. H. Whipple, F. S. Robscheit-Robbins, and L. L. Miller

University of Rochester, School of Medicine and Dentistry, Rochester, N. Y.

We propose to start with a brief outline of our concept of blood protein metabolism. Food proteins yield the amino acids absorbed from the intestinal tract, and the amino acids are synthesized, in the liver cells (and elsewhere), into plasma proteins. These plasma proteins (and amino acids) supply the protein requirements of the body cells. Normally, there is a considerable *reserve* of plasma protein-forming material (1 to 5 times the circulating mass), which reserve may be reduced by fasting, low protein diet, or plasma depletion. This depletion of protein reserves lowers the body resistance to infection and intoxication. These body protein stores, protein production, and protein wear and tear are in a nicely balanced or *steady state:* a dynamic equilibrium. These proteins can pass readily from plasma into cells, and the reverse, without loss of nitrogen. Whatever our concept of protein molecules and their passage through cell surfaces,[5] we must admit that protein does pass through cell surfaces readily, as a part of the normal protein metabolic exchange.

Hemoglobin in its production may derive in part from plasma protein, but hemoglobin contributes to the "protein pool" for exchange only when the red cell is destroyed, and then, after separation of the iron and complete loss of the pigment radicle, to the liver and bile. The body guards jealously the fabrication of hemoglobin and, given a real need for both plasma protein and hemoglobin (anemia and hypoproteinemia of double depletion), the protein flow favors hemoglobin. Under these circumstances, the hemoglobin is always produced in more abundance than the plasma protein, usually two or three to one (TABLE 1).

Globin from hemoglobin is broken down in the body continuously, due to red cell obsolescence, and is obviously perfectly suited to build new hemoglobin, with the addition of iron and the pigment radicle. These experiments[2] show that globin contributes to the "protein pool" material which serves well the protein requirements of the body. *Nitrogen balance* can be attained in dogs with *normal blood*, when abundant laked red cells are given intraperitoneally, and a basal ration

TABLE 1

INTRAPERITONEAL HEMOGLOBIN CONTRIBUTES TO PLASMA PROTEIN AND HEMOGLOBIN PRODUCTION IN ANEMIA AND HYPOPROTEINEMIA

Period 1 wk.	Weight	Protein intake		Protein output				Production ratio Plasma protein to hemoglobin	Total nitrogen	
		Type	Weekly	Hemoglobin		Plasma protein			Intake	Urinary output
				Level	Output per wk.	Level	Output per wk.			
	kg.		*gm*	*gm. per cent*	*gm.*	*gm. per cent*	*gm.*	*per cent*	*gm.*	*gm.*
		Dog 37-23		*Dog hemoglobin—intraperitoneal*						
2	20.9	Basal	20	6.7	19.5	4.7	11.0	56	3.2	10.9
3	19.4	Hb—31.9 gm.	30.7	9.4	57.3	5.4	25.8	45	5.2	8.2
4	19.2	Hb—42.9 gm.	40.8	8.1	48.2	4.8	21.7	45	6.9	8.3
5	18.9	Hb—44.0 gm.	42.2	9.4	12.9	4.5	5.8	45	7.1	8.7
6	17.9	Basal	15	8.2	25.5	4.3	11.1	44	2.4	6.2
7	17.0	Basal	11	8.2	1.6	4.4	0	—	1.8	6.3
		Dog 37-82		*Dog hemoglobin—intraperitoneal*						
1	17.1	Basal	18	7.7	16.2	4.6	9.5	59	2.9	8.4
2	16.4	Hb—38.6 gm.	37	10.9	12.9	4.5	5.8	45	6.2	7.1
3	15.9	Hb—60.7 gm.	58	10.7	45.5	4.7	19.7	43	9.8	6.8
4	15.3	Hb—44.2 gm.	51	11.7	18.3	4.3	7.3	40	8.6	6.4
5	14.7	Basal	9	11.7	1.7	3.4	0	—	1.4	5.3

supplies the needed carbohydrates, fats, minerals, and vitamin accessories. Globin is not as effectively used as is plasma protein, and there is a larger nitrogen output in the globin experiments, but these dogs remain close to nitrogen and weight balance. It is surprising that a relatively incomplete protein (globin) can contribute so effectively to the protein pool, as it must obviously be supplemented by the amino acids inadequately represented in globin, drawn from some reserve store, to produce cell protein or plasma protein. The whole globin in some dogs is more effectively used than the hemoglobin digests, suggesting that some of the globin may be used, without much breakdown and loss of nitrogen (TABLE 2).

Plasma proteins (in plasma), tested in the same manner by intraperitoneal injection, are completely utilized, with no loss of nitrogen, positive nitrogen balance, or weight balance, and no change in the albumin-globulin ratios.

Hemoglobin (globin) *digests* contribute effectively to body maintenance of nitrogen equilibrium. These digests are about as effective as

TABLE 2

PLASMA AND LAKED RED CELLS INTRAPERITONEALLY—NITROGEN BALANCE

Dog 43-141—Mongrel

Period No.	Blood proteins injected Total N	Total urinary N	Urea N+ NH_3-N	Total undetermined urinary N	Circulating plasma protein level	A/G ratio Tiselius	R.B.C. hematocrit	Weight
48 hrs.	*gm.*	*gm.*	*per cent*	*gm.*	*gm. per cent*		*vol. per cent*	*kg.*
1		3.31	76.7	0.77	5.72		41	9.8
2		2.55	73.4	0.68				
3		2.08	68.7	0.55				
4		2.34						
	Basal diet + whole blood plasma intraperitoneally							
5	4.60	1.80	75.5	0.44				9.4
6	4.17	2.53	83.5	0.42	7.64		38	
7	3.43	1.48	71.4	0.42				
8	1.86	1.30	65.8	0.45	8.90		45	9.3
9	3.80	1.61	67.4	0.52	9.70		40	
10	3.77	2.19	73.9	0.57				
11	2.07	2.70	69.9	0.82	9.98		41	9.4
12	3.81	2.52	72.3	0.70				
13	4.03	2.38	68.2	0.76				
14	4.06	2.14	69.3	0.66	9.32		35	9.5
Total..	35.6	20.7						
	Basal diet							
15		2.14	69.3	0.66		0.8	29	
16		2.00	63.7	0.73			30	
17		2.05	67.2	0.67	7.72		32	
18		1.69	66.0	0.57				
	Basal diet + laked red blood cells intraperitoneally							
19	4.31	2.00	61.5	0.77	6.64	1.0	36	9.2
20	4.62	2.56	72.0	0.72	6.61		43	9.2
21	4.10	2.74	70.6	0.80	6.50		47	
22	1.80	2.40	66.7	0 80				
23	4.09	2.06	71.5	0.59				
24	4.30	3.39			6.93		51	
25	2.25	2.46						9.1
26	3.62	3.34					46	
27	3.80	1.86				0.5		
28	4.63	2.79						9.1
)tal.	37.5	25.6						
	Basal diet							
29		2.88						
30		1.63						
31		1.60						9.0

whole hemoglobin in maintaining nitrogen balance, but cause a rise in undetermined nitrogen, not seen when hemoglobin alone is given intraperitoneally (TABLE 3).

TABLE 3*

LAKED DOG RED CELLS INTRAPERITONEALLY. NITROGEN RETENTION IMPROVED BY DL-METHIONINE, BUT NOT BY DL-ISOLEUCINE

Period (10 days)	Total N intake	Red Cell N injected	Total urinary nitrogen	At start of period		
				Weight	Circulating plasma protein level	R.B.C. hematocrit
	gm.	*gm.*	*gm.*	*kg.*	*gm. %*	*vol. %*
Dog 43-31						
1) Basal + Hemoglobin + dl-Isoleucine	19.49	16 80	25.47	14.8	5.22	66
2) Basal + Hemoglobin + dl-Methionine	19.48	16 45	16.78	14.0	5.56	64
3) Basal + Hemoglobin + dl-Methionine + dl-Isoleucine	20.91	16 75	16 73	14.3	5.75	63
Dog 43-346						
1) Basal + Hemoglobin	20.94	18.64	19 49	10.0	6.02	50
2) Basal + Hemoglobin + dl-Methionine + dl-Isoleucine	22 00	18.24	13 43	9.9	6 87	62
3) Basal + Hemoglobin + dl-Isoleucine	17.28	13 00	16.56	9.9	6.75	62

* TABLE 3 presents a summary of as yet unpublished data showing how nitrogen retention, during hemoglobin injection, is markedly improved in adult dogs by a small supplement of methionine, but not by isoleucine. This is, understandably, at variance with reports showing failure of immature rats to grow, on a dietary containing human or beef globin as the protein component

Pigment radicles derived from hemoglobin, given intraperitoneally, are thrown away and appear as surplus bile pigment, even when there is urgent need for all available nitrogenous material, given protein fasting, anemia, and hypoproteinemia in a bile fistula dog. The body, evidently, prefers to *make*, rather than conserve, the pyrrol aggregate (pigment radicle) (TABLE 4).

Given healthy dogs, fed abundant iron and protein-free or low protein diets, with sustained depletion due to bleeding, we can study the capacity of these animals to produce, simultaneously, new hemoglobin and plasma protein. The reserve stores of blood protein-producing

TABLE 4

BILE PIGMENT EXCRETION IN BILE FISTULA DOG INCREASED BY LAKED RED CELL INJECTIONS—ANEMIA AND HYPOPROTEINEMIA

Dog 40-41—Bile fistula

Period No	R. B. C. injected Total N	Total urinary N	Bile pigment excretion	Circulating plasma protein level	Hemoglobin	Weight
48 hrs	*gm.*	*gm.*	*mg.*	*gm. per cent*	*gm. per cent*	*kg*
Basal diet (0 86 gm. N per period)						
1		4 53	63	5.50	8.9	16 4
2		3.53	58			
3		2 73	68			
Basal diet + laked red blood cells intraperitoneally						
4	2.09	2.71	83			14.9
5	4.03	3.75	386		12 7	
6	2.54	4.97	525	5.50		
7	3.66	3.82	459			
8	2.26	3.30	222			14.1
9	4.63	3.07	391			
Total . .	19.2	21.6				
Basal diet						
10		3.51	192	4.83	15.4	13.6
11		3.19	104			
12		1.61	96			

materials, in this way, are largely depleted, and levels of 6 to 8 gm. per cent for hemoglobin, and 4 to 5 gm. per cent for plasma protein can be maintained, for considerable periods of time. These dogs are very susceptible to infection, and to injury by many poisons. Under such conditions, these *anemic and hypoproteinemic* dogs will use very efficiently a variety of *digests* (serum, hemoglobin, and casein) and the growth mixture (Rose) of pure *amino acids*. Nitrogen balance is maintained, and considerable new blood proteins are produced[3] (TABLES 5 and 6).

Amino acid mixtures are of especial interest. The growth mixture of 10 amino acids (Rose) is well utilized by mouth, subcutaneously, intraperitoneally, or by vein. These ten essential amino acids are threonine, valine, leucine, isoleucine, lysine, tryptophane, phenylalanine, methionine, histidine, and arginine. Glycine is usually added to this mixture. By mouth, the amino acids are utilized a little more completely than when given parenterally. Our experiments give no evi-

TABLE 5

HEMOGLOBIN DIGESTS INTRAVENOUSLY—NITROGEN BALANCE
HEMOGLOBIN AND PLASMA PROTEIN LEVELS

Dog 39-251

Period No.	Hemoglobin digests injected Total N	Total urinary N	Urea N + NH_3-N	Total undetermined urinary N	Circulating plasma protein level	R.B.C. hematocrit	Weight
18 hrs.	*gm.*	*gm.*	*per cent*	*gm.*	*gm. per cent*	*vol. per cent*	*kg.*
Basal diet (0.50 gm. N per period)							
1		1 48	59.5	0 60			
2		1 78	55.5	0 80	5 56	49	10.9
Basal diet + hemoglobin digest intravenously							
3	4 36	5 10	59.5	2 07			
4	4 36	3 55	60 3	1 41			10 8
5	4 36	5 33	66 8	1 77			
6	4 36	5 28	65 5	1 72			
7	4 36	4 08	70 5	1 20			
8	4 36	4 53	72.1	1 26			11 1
9	4 36	4 50	67 8	1 45	5 37	46	
Total.	30 5	32 4		10 88			
Basal diet							
10		1 92	64 1	0 69			
11		1 76	60 4	0 70			11 0

TABLE 6

PRODUCTION OF HEMOGLOBIN AND PLASMA PROTEIN DUE TO A CASEIN DIGEST

Period 1 wk.	Weight	Protein intake: Type	Protein intake: Weekly	Protein output: Hemoglobin: Level	Protein output: Hemoglobin: Output per wk.	Protein output: Plasma protein: Level	Protein output: Plasma protein: Output per wk.	Production ratio plasma protein to hemoglobin	Ratio protein output to intake	Total nitrogen: Intake	Total nitrogen: Urinary output
Dog No. 40-155. Casein Digest P36092											
	kg.		*gm.*	*gm. per cent*	*gm.*	*gm. per cent*	*gm.*	*per cent*	*per cent*	*gm.*	*gm.*
1	15.1	Digest—vein	182	5.1	41.1	6 6	20.0	51	—	—	25.1
2	14 8	Digest—vein	188	6.9	20.5	6.5	16.7	84	—	59 2*	25.2
3	14.0	Basal	26	6.8	10.0	6.0	8.2	85	—	—	14.7
4	13.3	Basal	30	5.9	10 7	5 9	7.4	72	25	8 8*	10.7
Total output net. . . .					62		45	73			
1	14 1	Digest—oral	144	6 3	27.5	5 0	17.8	67	—	—	—
2	13.7	Digest—oral	136	6 3	1 3	5 5	—	—	—	—	—
3	13.8	Digest—oral	131	7 7	13.9	5.3	8 1	61	—	—	—
4	13.5	Basal	18	6 2	22.5	4 8	12 2	57	25	—	—
Total output net. . .					73		38	52			

dence that the unnatural isomers of the amino acids are toxic; some are probably used in the body. These amino acid mixtures can be given rapidly, in 10 per cent solution, parenterally, and cause less clinical disturbance than any protein digests so far tested. *Glutamic acid,* in digests or amino acid mixtures, is not well tolerated by vein and may induce vomiting. Abundant production of plasma proteins, in standardized dogs, is readily demonstrated in experiments in which these amino acids are the sole source of nitrogen. The production of new plasma protein, due to amino acids, in general corresponds favorably with the response to high grade diet protein in equivalent amounts (TABLE 7). The first observations were upon dogs, but subsequent

TABLE 7

PRODUCTION OF HEMOGLOBIN AND PLASMA PROTEIN DUE TO AMINO ACIDS NECESSARY FOR GROWTH (ROSE)

	Protein intake			Protein output				Production ratio plasma protein to hemoglobin	Ratio protein output to intake	Total nitrogen	
				Hemoglobin		Plasma protein					
		Type	Weekly	Level	Output per wk.	Level	Output per wk.			Intake	Urinary output
	kg.		*gm.*	*gm. per cent*	*gm.*	*gm. per cent*	*gm.*	*per cent*	*per cent*	*gm.*	*gm.*
		Dog No. 40-32. Amino acid mixture Vc daily									
1	13.9	Basal	34	6.8	1.7	4.7	0	—	—	—	—
2	13.8	Amino acid—oral	177	8.8	18.5	4.8	12.4	67	—	28 3	15 8
3	12.9	Amino acid—oral	177	7 5	38.0	4.9	19.8	52	—	28 3	15.8
4	12.8	Amino acid—vein	177	8 8	15.8	4.7	7 7	49	23	28 3	16.7
5	12 6	Basal	0	8 8	1.6	3.9	0	—	—	0	4 03
	Total output net.....				89		34				
		Dog No. 39-2. Amino acid mixture Vc daily									
1	14.1	Basal	40	5 4	7.6	4.6	4 8	63	—	—	—
2	13.5	Amino acid—oral	100	7 3	1.7	4.6	0	0	—	—	—
3	13.2	Amino acid—oral	77	9 8	12 3	4.3	6 6	54	32	28 3*	16 7*
4	12.6	Basal	0	9.8	2.1	4.3	0	0	—	—	—
5	12.3	Basal	0	11.0	2.1	4.0	0	0	—	0	15 2*

* Combined figures for two periods.

observations on human patients gave the same picture.[1] These amino acid mixtures are well utilized by patients suffering from chronic infection (colitis), gastrointestinal disturbances (partial obstruction), or cancer cachexia.

The *maximal output ceiling* for hemoglobin, in anemia due to blood loss, is about 60 gm. per week, the dog receiving a rich protein diet, plus high iron intake. Ferrous and ferric salts are equally effective. Iron, intravenously absorbed, plus a rich protein diet, may push this level up to 90 to 100 gm., per week. Evidently, iron absorption is a limiting factor in these experiments.

Maximal output for hemoglobin, plus plasma protein, in doubly depleted dogs,[4] may reach 120 to 130 gm. per week and, using intravenous iron, may reach 140 to 160 gm. per week (TABLE 8).

TABLE 8

MAXIMAL BLOOD PROTEIN PRODUCTION—GRAMS PER WEEK
ANEMIA AND HYPOPROTEINEMIA

Period 1 wk.	Weight	Protein intake		Protein output, weekly					Ratio protein output to intake	Total nitrogen, gm.	
				Hemoglobin		Plasma Protein		Total output			
		Type	Weekly	Level	Output	Level	Output			Intake	Urinary output
	kg.		*gm.*	*gm. per cent*	*gm.*	*gm. per cent*	*gm.*	*gm.*	*per cent*	*gm.*	*gm.*
					Dog 40-33						
1	18.5	Basal	19	8.6	26.3	3.9	9.0	35.3		3	9.3
2	18.6	Liver, beef, iron	973	7.8	89.0	5.1	44.9	133.9	14	157	37.4
3	18.8	Liver, beef, iron	1037	7.8	38.3	5.5	23.5	61.8	6	166	47.7
4	20.6	Liver, beef, iron	1037	9.5	69.4	5.4	38.6	108.0	10	166	57.0
5	20.2	Liver, beef, iron	1037	6.6	74.3	5.3	44.7	119.0	11	166	75.7
Average output per wk.					67.8		37.9	105.7			
6	20.7	Basal	19	6.6	23.3	4.6	10.7	34.0		3	18.3
7	18.3	Basal	19	10.3	17.9	3.8	6.6	24.5		3	10.1
					Dog 37-85						
1	15.2	Basal	84	6.9	12.3	4.9	8.0	20.3		13	8.4
2	16.4	Liver, beef, iron	764	11.8	47.8	5.3	26.7	74.5	10	123	37.6
3	17.4	Liver, beef, iron	764	12.3	57.1	5.5	25.3	82.4	11	123	50.8
4	17.5	Liver, beef, iron	764	12.8	78.5	5.6	42.9	121.4	16	123	68 1
Average output per wk.					61.1		31 6	92.8			

Maximal output for plasma protein, alone, in hypoproteinemia due to plasmapheresis, reaches 60 to 70 gm., per week, but *this is not the true ceiling*. Technically, we cannot remove the new plasma protein as fast as it is formed, and the hypoproteinemia is not maintained, in

the face of a rich protein diet intake. Furthermore, the evidence points to the protein circulating pool as contributing to the accretion of tissue protein, in such dogs with a strong positive nitrogen balance and weight gain.

Maximal figures for *hemoglobin production* in anemia run close to 1 gm. hemoglobin per kilo per day. Maximal figures for *new hemoglobin plus plasma protein* production, in simultaneous anemia and hypoproteinemia, using iron given intravenously, may reach 1.5 gm. blood protein per kilo per day. The actual, maximal *plasma protein production* equals about 1 gm. per kilo per day, but the true production ceiling cannot be reached by this technique (TABLE 9).

TABLE 9

MAXIMAL HEMOGLOBIN PRODUCTION—GRAMS PER WEEK
STANDARD CONTINUING ANEMIA OF 6 TO 8 GM. HEMOGLOBIN

Dog No.	Dog Average normal weight kg.	Daily diet Liver 300 gm.	Daily Diet Salmon bread plus iron 400–450 mg.			Daily diet Salmon bread, liver plus iron 400–450 mg.			Daily diet Liver, salmon bread plus iron by vein 24 mg	
			Reduced	Ferric	Ferrous	Reduced	Ferric	Ferrous	Colloidal Fe	Estimated plasma protein removed
39-1	18 0	45*	60	51	50	62	58	66	92	71
40-26	14.5	50	59	45	60	56	63	—	80	49
37-21	18.0	41*	58	64	54	53	52	51	82	53
34-148	18.0	44*	49	49	54	47	63	71	106	58
34-145	20.0	47*	47	55	57	63	82	56	95	52
37-89	14 0	40*	32	54	42	39	48	47	84	47
33-14	12.0	38*	28	41	44	38	58	56	75	48
Average....	16 3	44	48	51	52	51	61	58	88	54

Average 2 to 6 experiments.

REFERENCES

1. **Bassett, S. H., R. R. Woods, F. W. Shull, & S. C. Madden**
 1944. New Eng. J. Med. **230**: 106.
2. **Miller, L. L., F. S. Robscheit-Robbins, & G. H. Whipple**
 1945. J. Exp. Med. **81**: 405.
3. **Robscheit-Robbins, F. S., L. L. Miller, & G. H. Whipple**
 1943. J. Exp. Med. **77**: 375.
4. **Robscheit-Robbins, F. S., L. L. Miller, & G. H. Whipple**
 1945. J. Exp. Med. **82**: 311.
5. **Whipple, G. H.**
 1942. Am. J. Med. Sci. **203**: 477.

DISCUSSION OF THE PAPER

Dr. S. H. Bassett (*University of Rochester, School of Medicine and Dentistry, Rochester, N. Y.*):

The interesting experiments reported by Dr. Robbins and Dr. Miller suggest that properly prepared protein hydrolysates, or suitable mixtures of pure amino acids, may be utilized as well as, or better than, plasma proteins or globin. Dr. S. C. Madden and I have been particularly interested in the parenteral use of such mixtures and have found them well tolerated, even in very ill patients and those whose stores of protein were greatly depleted. Unfortunately, the depleted human, especially when suffering from obstruction to the gastrointestinal tract, presents a very difficult therapeutic problem. It is not within my province to discuss this matter in detail, since it is to be considered by other contributors to this symposium. I would like to mention, however, that one of the limiting factors in the utilization of the nitrogen given, seems to be the caloric intake. When all feedings were parenteral, it was difficult to provide more than 1400 to 1600 calories a day as amino acids and glucose. Under these circumstances, most of the amino acids were promptly deaminized and appeared in the urine as urea and ammonia, probably because they were used as a readily available source of energy. While such a program may be beneficial in maintaining the *status quo*, rapid repletion is noted only when the caloric intake rises to 2500, or more, per day. If full benefit from parenteral feeding is to be obtained, it seems urgent to find a means of administering non-protein calories at a considerably higher level than is, at present, feasible.

NITROGEN METABOLISM IN ACUTE AND CHRONIC DISEASE

By John P. Peters

Yale University, School of Medicine, New Haven, Connecticut

The muscular wastage that accompanies febrile infections has been long recognized. As much as 60 years ago (mentioned by Shaffer and Coleman[1]), it was demonstrated by Friedrich Müller that patients with typhoid fever lost comparatively large quantities of nitrogen in the urine. The subject was examined in detail, by Shaffer and Coleman,[1] in 1909; by Kocher,[2] in 1914; and by DuBois and associates,[3-6] from 1915 to 1922. The term applied to these losses of nitrogen, "toxic destruction of protein," though based on conceptions that have been largely abandoned, has been so long employed that it may be retained, at least until enough has been learned of the nature of the reactions involved to justify the application of a more appropriate term. Subsequently, it has been shown that similar destruction of protein is encountered, after acute hemorrhage,[7, 8] operations and injuries,[9-12] and a variety of other conditions.

The disorder, which had been a matter of desultory interest, became a subject of urgent inquiry, during the war, because of the necessity for hastening the return of the sick and wounded to active duty. With the discovery of new agents, which have greatly reduced mortality, complications, and permanent disability, chiefly by limiting or eliminating the effects of infections, it was assumed that the duration of disability would be proportionally reduced. Actually, however, this did not prove to be the case. Among the factors which appeared to retard rehabilitation and convalescence, loss of strength and nutrition were prominent. This was particularly unexpected, since the diets provided for sick and wounded in the armed forces were, according to accepted standards, more than adequate for their purposes. The reports that have come out of these renewed inquiries are conflicting in the extreme. Most observers have been disappointed in their efforts to prevent losses of nitrogen, after severe injuries and operations, by dietary measures.[13, 14, 15] Some have claimed that, if a sufficient excess of protein, or its equivalent in the form of protein hydrolysates or amino acids, is given, positive nitrogen balances can be achieved.[11, 16, 17, 18] When honest men disagree, the causes of their differences are likely to be found in the condi-

tions of their experiments, rather than in the accuracy of their data. It may be well, therefore, to examine in some detail what is known of the nature of the "toxic destruction of protein."

It was first generally believed that this phenomenon was merely an expression of accelerated oxidations associated with fever. This has been disproved, in a variety of ways. Kocher[2] showed that a normal person, subsisting on a diet containing minimal amounts of protein, but high calories, could perform heavy muscular work without any increase of urinary nitrogen. On the other hand, it was impossible, by similar diets, to minimize the nitrogen excretion of patients with febrile diseases. DuBois and his associates[3-6] were unable to prevent losses of nitrogen, in similar conditions, by the administration of diets containing adequate quantities of protein and calories far in excess of the needs of the patients, which were measured by both direct and indirect calorimetry. Neither accelerated oxidations nor fever can explain the uneconomical expenditure of protein, after fractures,[13] and during convalescence from acute infections.[15]

The degree and duration of the destruction of protein seem to vary with the severity of the injury or disease. This is evident in the studies of Kocher,[2] of DuBois and his associates,[3-6] and others who have investigated the problem.[12, 15] The term, *severity,* however, is somewhat undefinable. As yet, it can be measured in no objective terms. Grossman *et al.*[15] were able to prevent nitrogen losses, without difficulty, after repair of inguinal hernias, but were quite unsuccessful after appendectomies. After major fractures, negative nitrogen balances were regularly observed, but, after osteotomies, nitrogen equilibrium was easily established.[14, 15] In the case of infections, nitrogen losses appear to be related to the inherent gravity of the disease, not to its clinical manifestations. In the studies of Grossman and associates,[15] negative balances persisted in meningococcus meningitis, after the temperature had become normal and symptoms and signs of the disease had been abolished by drug therapy. In this disease and in scarlet fever, treated by sulfonamide drugs, negative nitrogen balances frequently continued, despite high protein diets, until patients were discharged from the hospital, apparently well. This may explain why these drugs, in spite of their dramatic effects upon the overt signs and symptoms of these diseases, have not proportionately hastened convalescence and rehabilitation.

"Toxic destruction of protein" is a phenomenon characteristic of acute injury to previously healthy individuals. It is a self-terminative reac-

tion. If the pathological process continues long enough, the loss of nitrogen gradually diminishes, then disappears. Finally, provided that an adequate diet is given, a positive nitrogen balance can be established, though the disease or injury continues. This is evident in the early experiences of Shaffer and Coleman,[1] and Coleman and DuBois,[9] with typhoid fever. In three of their cases, they established positive nitrogen balances during relapses, although they had failed in their efforts during the primary phases of the disease. Presumably for similar reasons, little difficulty is encountered, provided a patient can be made to take food, in preventing wastage of protein during tuberculosis[20] and other chronic infections.[15] Browne and his associates[14] have named the period in which protein can be stored, the *anabolic phase* of injury, while, to the early period characterized by nitrogen losses, they apply the term, *catabolic phase*. They have found that renewed injury during the anabolic phase may result in little or no waste of protein. The duration of the catabolic phase varies, again, with that incommensurable quality, the gravity of the injury or infection.

The incidence and duration of toxic destruction of protein is, therefore, not altogether predictable. It follows that it is extremely difficult to control and to interpret studies of the phenomenon. These difficulties are well illustrated in the studies of typhoid fever by Shaffer and Coleman.[1] It is impossible, in most instances, to decide whether a change of protein metabolism is referable to dietary regulation or to the course of the disease. All claims that the "toxic destruction of protein" has been mitigated or prevented by dietary treatment must be analyzed, with full consideration of these conditioning features. Madden and Clay[18] were able, in some instances, by giving large quantities of mixtures of amino acids or hydrolysates of protein, to produce positive nitrogen balances, in dogs with sterile abscesses induced by injections of turpentine. Such lesions are, however, of short duration. Hirshfeld and his associates,[16] by giving 150 or more grams of protein and over 4000 calories, daily, to certain burned patients, succeeded in establishing positive nitrogen balances, almost from the outset of their treatment. In other patients, similar diets were not effective. This is true, also, of the series of cases presented by Elman[17] and by Brunschwig *et al.*[10] In Elman's series, some of the patients were, obviously, chronically ill and wasted. In all cases, administration of amigen was instituted only after the lapse of 4 to 6 days. This is the period in which Brunschwig[10] and others noted the greatest postoperative wastage. Nevertheless, Elman did not succeed in establishing

positive nitrogen balances in all his cases. The successes of Mulholland, Co Tui *et al.*[11] in sparing protein after gastrectomy may have been referable to the poor initial condition of their subjects. These operations were performed, chiefly, on patients with chronic wasting disease. In many reports, treatment and measurements of nitrogen metabolism have not been instituted until some time after the initial injury was sustained.

Another confusing feature is introduced, when patients are losing protein from exudative lesions, *e.g.*, in burns. In this case, the nitrogen exchange cannot be measured by analyses of food, urine, and feces, alone. Unless the losses of nitrogen in the exudate are also measured, the impression is gained that the nitrogen balance is more favorable than it actually is. This criticism applies to the studies of burned persons, reported by Cope and associates,[19] in which positive balances were claimed.

There is some evidence that, when a specific deficiency of serum protein or hemoglobin is produced by plasmapheresis, transudation, or hemorrhage, direct replacement of these materials, by transfusions of plasma or whole blood, may be effective in correcting the specific deficit. Hirshfeld *et al.*[16] have rightly pointed out that nitrogen retained for these purposes must not be credited to general protein metabolism. There is evidence that the proteins of blood and serum are not altogether adequate for general nutritive purposes, because they are deficient with respect to certain essential amino acids, especially isoleucine.

To date, no one appears to have succeeded in preventing nitrogen losses throughout the acute or catabolic phase of the reaction to injuries of all kinds occurring in previously healthy, well-nourished individuals. It still remains an open question whether such success can be achieved by dietetic measures. Certainly, the protein stores of the body cannot be preserved in the face of such injuries, unless dietary treatment is given far more vigorous attention than it has received in the past. If this disorder is an inherent part of the general process of repair, efforts to correct it may be misdirected; the old principles of starvation and light diets may have been sound. Hirshfeld, Abbott, *et al.*[16] did, indeed, find that the early administration of large amounts of protein to burned patients provoked nausea, vomiting, diarrhea, and untoward symptoms. These were evident only in the initial stages, immediately after injury. After the first few days, however, these symptoms disappeared, and patients were able to take large quantities

of protein, without deleterious effects. Since "toxic destruction of protein" is a self-terminative reaction of unpredictable duration, the advent of the anabolic phase should be anticipated by the provision of generous diets before the catabolic phase has ended. There is an impression that such diets hasten the appearance of the anabolic phase, and there can be little doubt that, after this phase has begun, they accelerate restoration of the depleted supply of protein in the body.

Certain other features of "toxic destruction of protein" and its response to treatment deserve mention. The very inception of this reaction in human subjects usually escapes observation. For this part of the story, we are dependent upon animal experiments or a study of the effects of operation. Cuthbertson,[9] who first approached the surgical aspects of the problem, found that nitrogen excretion, immediately after operation, was not great. Only after the lapse of 3 or 4 days did it become considerable. This was confirmed by Grossman *et al.*[13] Howard[12] has made somewhat similar observations, after fractures. This initial economy may be more apparent than real. It may arise only from the fact that patients take little or no nourishment, immediately after operation. The administration of amigen, in amounts equivalent to 75 gm. of protein per day, on the first two postoperative days, by Grossman,[13] did not significantly diminish negative nitrogen balances. The extra nitrogen from the amigen solution was, apparently, wasted in the urine. Reports of the amounts of nitrogen wasted in subsequent days vary greatly. For this variability, again, the great differences in dietary management may be largely responsible. Although Shaffer and Coleman[1] were not successful, in most instances, in preventing losses of nitrogen, during the early stages of typhoid fever, they did show, in several instances, that the degree of wastage could be influenced by the caloric values of the diets given. Nitrogen losses rose distinctly when the diets contained less than 2000 calories daily, and diminished when they contained 3500 calories, or more. This does not mean that these losses were proportional to the calories taken; but only that, when the caloric intake fell below the daily energy demands, extra protein was expended. In only one case, in which attempts at feeding were ineffectual, were the negative balances extremely large. In the others, they were seldom as great as 12 gm. daily, even at the height of the fever. Brunschwig *et al.*[10] have also emphasized the need for calories. The losses in Kocher's[2] patients, who were given high calory diets with minimal protein, seem large; but they must be compared with the nitrogen excreted by healthy persons, on similar diets. On this basis, the urinary nitrogen that can rightly be attributed to the

effects of the infections are not extremely large and might have been further reduced, if it had been feasible for the patients to take as large diets as the normal subjects took.

This is a point of some importance. The impression has frequently been given that the injured animal has a malicious tendency to destroy protein. Actually, it seems rather to be under compulsion to waste a certain amount of its own tissue protein. Over and above this, in the utilization of protein, it follows the same principles as the normal animal. It will waste more protein, if it is not provided with sufficient calories in another form. It has no capacity to store large amounts of protein.

Whether, by the administration of unusually large quantities of protein, as well as calories, this waste of native protein can be prevented or not, is still an open question. There is no clear evidence, in the cases of Shaffer and Coleman,[1] that variation of dietary protein has an important effect. However, their ability to test this phase of the subject was limited by tradition and by the appetites of their patients. With the aid of protein hydrolysates and the elimination of fears that large quantities of protein are harmful, it has been possible to test the effects of almost unlimited amounts of protein. It would be surprising, indeed, if this had no effect on nitrogen balances. It is clearly established that its effect is relatively trivial, that protein is used extravagantly in the acute phases of injury.

"Toxic destruction of protein" has been ascribed, as its name implies, to autolysis of tissue, either at the site of injury or throughout the body. It is observed, however, in certain cases in which there is no good evidence of such destruction. There is no evidence that tissue destruction was proceeding at a rapid pace in the patients convalescing from meningitis and scarlet fever, which were mentioned above.[15] After fractures, nitrogen wastage continues after reparative processes are well advanced.[16] Furthermore, there is no reason to assume that products of autolysis occurring within the body should be less suitable, for synthetic processes, than products of digestion or autolysis conducted in the gastrointestinal tract or the test tube.

The nitrogen balance depends, not only upon the quantity, but also upon the quality of the protein in the diet. It has been demonstrated, especially by Rose,[21] that, of the total number of known amino acids, 10 cannot be synthesized by animals, but must be supplied in the food, in proper quantities and proportions. If supplied such a mixture, animals will survive and grow, with no other source of protein.[22, 23] Such

mixtures of amino acids are utilized as efficiently, when given by vein, as they are when given by mouth. Digests or hydrolysates of efficient proteins have been devised which, when given orally or intravenously, will also support life. If a protein or hydrolysate contains insufficient amounts of one or more essential amino acids, it will not support life and growth. If given to an animal, in lieu of an efficient protein, such an incomplete hydrolysate is not retained; the nitrogen balance of the animal becomes negative. The body can not utilize incomplete proteins. These essential amino acids are required, not only for the formation of protein, but also for the formation of other nitrogenous compounds. It is conceivable that, under certain conditions, a specific need might be created for extra quantities of a particular amino acid to provide extra amounts of some one of these compounds.

It has been suggested that such a state is produced by acute injury or disease. Indeed, Croft and Peters,[24] by administering methionine to rats, diminished the nitrogen losses after thermal burns. If, however, "toxic destruction of protein" is referable to the need for an excess of some particular amino acid, it should respond more adequately to the administration of large quantities of protein containing this acid. Little evidence that protein metabolism is being diverted into unusual channels is found in the urinary nitrogen partition. The usual proportion of nitrogen is excreted as urea + ammonia, and extra dietary nitrogen appears in the urine almost entirely in these forms. This is evident from FIGURE 1, in which the urinary nitrogen, other than urea + ammonia, from two patients, is plotted against total nitrogen metabolism and against nitrogen balances. No significant correlation can be found in either figure. In the data of Shaffer and Coleman,[1] and of Kocher,[2] nonurea + ammonia nitrogen is somewhat higher, in the earlier stages of infection, than it is later. It does not, however, in the longer studies of Shaffer and Coleman, bear any consistent relation to either nitrogen metabolism or nitrogen balances. The excretion of creatinine and uric acid, in these same studies, is distinctly greater than normal, in the early stages of infections, but this, again, cannot be correlated with nitrogen metabolism or nitrogen balances. The incidence of creatinuria, in the same experiments, is somewhat irregular. In our own studies, urinary creatinine and creatine[25] are quite as unrelated to the "toxic destruction of protein" *per se*. Evidently, whatever may be the disturbances of the intermediary metabolism of protein, the major proportion of its products is excreted in the most highly oxidized forms.

Among the demonstrable results of insufficient dietary protein, reduction of the concentration of albumin in the serum is generally recognized.[26, 27, 28] So clearly has the relation of serum albumin to the protein supplies in the body been established, that hypoalbuminemia has

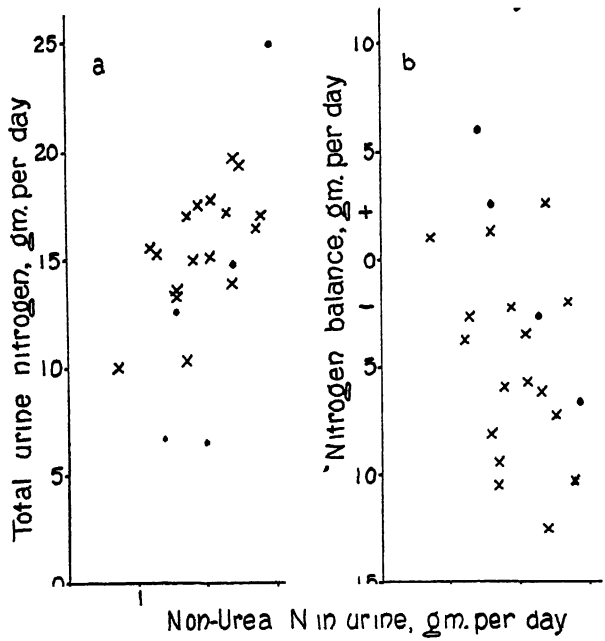


FIGURE 1 The relation of non-urea + ammonia N of urine to. (a) total nitrogen of urine, (b) nitrogen balance. Circles represent one subject, crosses another.

been generally accepted as a diagnostic sign of protein deficiency. During the catabolic phase of the reaction to injury, serum albumin is usually distinctly reduced. This deficiency has, generally and quite naturally, been attributed to malnutrition. In this case, serum albumin should decline further, as nitrogen wastage progressed. From certain casual observations in FIGURE 2, however, no such relation is evident. Albumin sometimes rises appreciably, in the face of large negative nitrogen balances, in some instances greater than 100 gm., which corresponds to the protein, in from 2.5 to 3.0 kgm. of muscle. It cannot be inferred, from these data, that loss of tissue protein does not cause serum albumin to decline. This has been incontrovertibly established. The only possible conclusion is that the initial reduction of serum albumin is, in this case, referable to some other factor. This view is supported by the fact that serum albumin was usually subnormal, at the time of the initial examinations.

Unless our concepts of the intermediary metabolism of protein are erroneous, both the products passing to the tissues for synthesis and those in process of transportation from the tissues to the liver and kidneys, for conversion to the excretory products, urea and ammonia,

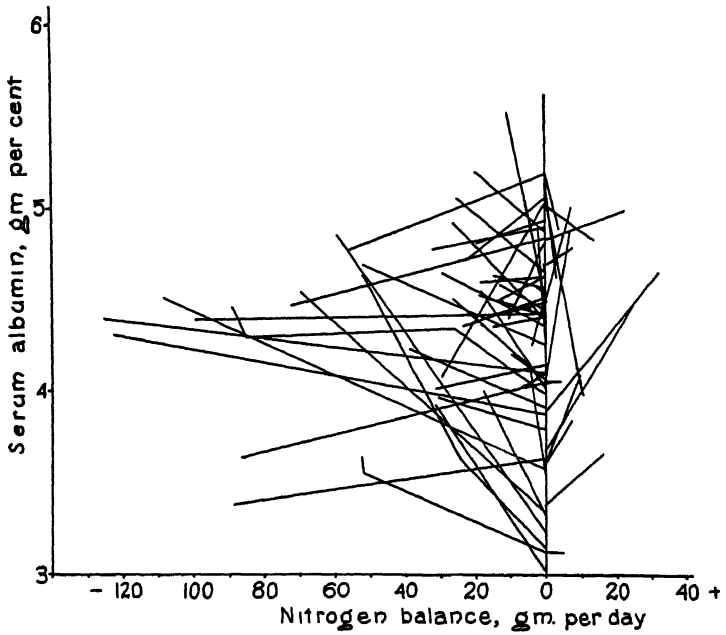


FIGURE 2. Comparison between nitrogen balances and changes of serum albumin. Blood for measurement of serum albumin was taken, in the postabsorptive state, at the beginning and end of each study, and occasionally at intervening times. The vertical line represents, therefore, in most instances, the beginning of a study.

are included in measurements of the concentration of amino acid nitrogen in the blood plasma. It was, therefore, anticipated that this fraction of the nonprotein nitrogen might be elevated during the catabolic phase of the injury reaction. When Man[29] analyzed the plasma of patients in this condition, however, she found that the amino acid nitrogen was seldom high and often distinctly low, even in the face of high dietary protein and large negative nitrogen balances. This is illustrated in FIGURE 3, in which plasma amino acid nitrogen is plotted against total nitrogen metabolism and nitrogen balances.

All these observations suggest that injury effects a sudden and profound transformation of the metabolism, that manifests itself, not only in the uneconomical utilization of protein, but also in certain demonstrable changes in the chemical composition of the blood. This has led us to a more detailed study of the immediate reactions of patients

to injury, in the hope that this may throw some light on the nature of the subsequent reaction. For this purpose, surgical operations were chosen, in order that observations could be made, previous to the injuries. First of all, it was ascertained that urinary nitrogen excretion,

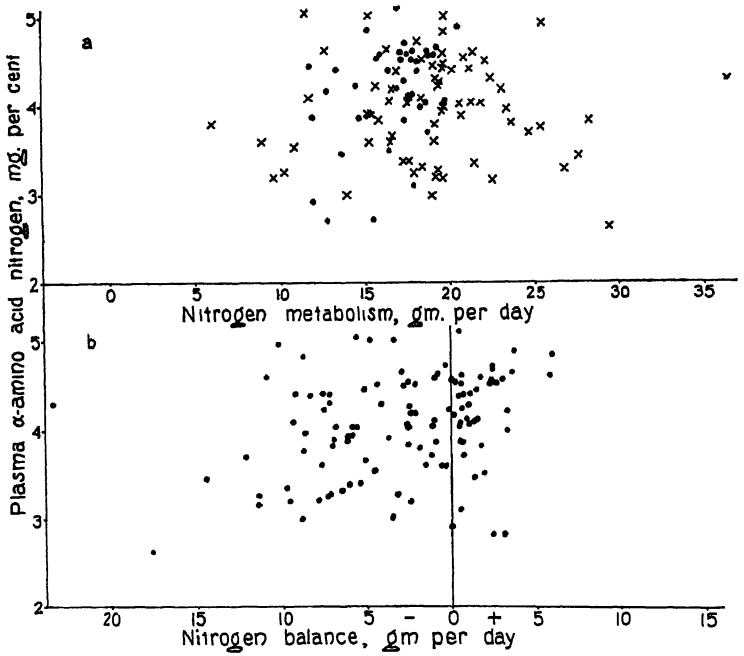


FIGURE 3. (a) The postabsorptive concentration of α-amino acids of plasma, compared with nitrogen metabolism (nitrogen intake or nitrogen output, whichever was the larger), and (b) nitrogen balance. In (a), circles = positive, crosses = negative, balances.

on the day of operation and the following day, was not large, as Cuthbertson,[9] Howard,[12] and others[15] had already noted. This is evident from TABLE 1, in which it is also shown that blood nonprotein nitrogen does not usually rise. Urinary total nitrogen and creatinine were usually somewhat greater, on the second, than on the first day, but still moderate. This tendency to increase can be ascribed, in part, to the difficulty experienced by many of the patients in emptying their bladders at the proper time, the first morning after operation.

It was next discovered that plasma amino acid nitrogen fell sharply, immediately after operation. This is illustrated in TABLE 2. The drop is roughly related to the initial concentration of amino acids, which, in turn, reflects the antecedent condition of the patients. Amino acid nitrogen was highest in patients, previously in good health, who were

TABLE 1

NITROGEN METABOLISM ON THE DAY OF AND THE DAY FOLLOWING OPERATION

Case No.	Nonprotein nitrogen mg. per cent —Day*— 0	1	Urine nitrogen gm. per day —Day*— 0	1	Case No.	Nonprotein nitrogen mg. per cent —Day*— 0	1	Urine nitrogen gm. per day —Day*— 0	1
A83628	35	34	8.6	12 6	B68027	32	32	7.3	11.2
A44427	43	43	8.7	7.8	A79194	26	25	5.0	7.3
B44458	28	25	8 9	10 8	B62484	30	26	9.5	9.1
B61691	33	33	10.4	17 9	B49302	29	26	7.0	16 8
B53964	41	50	4.7	7 8	B59857	30	29	10 7	13 5
B8104	28	27	8 0	16.3	61724	32	31	16 8	18.5
A44648	34	34	6 0	9.7	9093	36	38	5 5	8.9
B62115	26	46	4.8	13.5	B68658	31	25	5 0	9.0
34340	32	25	6.4	9.4	B69576	30	30	8.1	10.7
A11178	30	25	1 3	8 1	B5775	28	28	6 6	10.7
90248	25	30	5 8	13 5	B68445	31	18	2 7	7.7
B69089	28	29		11 9	34038	26	24	5.2	9.6
25165	31	31	4.9	10 6	B61943	34	22	5.9	5.3
B68338	29	22	5 9	8 0	A94807	34	37	5.6	10 3
A10607	31	29	9 4	14 0	B68886	22	25	5 0	9 0

* 0 = the day of operation, 1 = the day following operation.

subjected to the gravest operations. The decline was also correlated with the severity of the operation. Amino acid nitrogen fell furthest, therefore, in patients previously in good health, who were subjected to

TABLE 2

THE EFFECT OF OPERATION ON PLASMA AMINO ACID NITROGEN

Case No.	Plasma amino acid nitrogen, mg. per cent —Postoperative day* 0	1	(ff)	Case No.	Plasma amino acid nitrogen, mg. per cent —Postoperative day* 0	1	(ff)
A83628	4.95	3 45	4.17 (6)	B68027	3 97	3 73	4 12 (2)
A44427	4.85	3 54	4 17 (6)	A79194	3 96	3 42	3 51 (7)
B44458	4.54	4.48	4 36 (2)	B62484	3 94	3.31	3 87 (2)
B61691	4 53	4.13	4 32 (13)	B49302	3.90	3 27	3 33 (2)
B53965	4 44	3.97	3 57 (4)	B59857	3.85	3.03	3 16 (2)
B8104	4.44	3.24	4 81 (7)	61724	3.84	2.94	4.61 (5)
A44648	4.39	3.81	3 98 (4)	9093	3.73	3.41	
B62115	4.34	4.09	2 93 (2)	B68658	3.72	3.54	3 74 (8)
34340	4.30	2.71	4.08 (7)	B69576	3.70	3.38	
A11178	4.24	3.20	3.75 (7)	B5775	3 60	2.93	3.31 (11)
90248	4.24	3.74	4.57 (2)	B68445	3.51	2.73	4.23 (4)
B69089	4.17	3.14	3 99 (9)	34038	3.27	3.56	4.41 (2)
25165	4.16	3.64	4.22 (8)	B61943	3.19	2.85	3.74 (9)
B68338	4.16	3.00	2 58 (2)	A94807	2.75	3.04	2.93 (5)
A10607	4.14	3.57	3.93 (3)	B68886	2.51	2.75	2.49 (2)

0 = the day of operation, 1 = the day following operation. Figures under (ff) represent subsequent determinations, at intervals after operation indicated by the figures in parentheses.

the gravest operations. The duration of the amino acid deficiency depended upon the gravity of the operation and the postoperative course. Restoration was delayed longest in those patients who had the most serious operations or postoperative complications.

In another series of cases, it has been demonstrated that serum albumin behaves in a similar manner, declining precipitately, in the first 48 hours following operation. TABLE 3 includes only the first 4 cases

TABLE 3

THE EFFECT OF OPERATION ON SERUM ALBUMIN

Case No.	Serum albumin per cent				
	*Postoperative day**				
	0	1	2	(ff)	
B50516	5 10	4.03	3 72	3 24	(7)
B72371	3 79	3.53	2 16	3 31	(8)
B58730	4 50	4 62	2 65		
B5775	2 95	2 55	2 55		

* 0 = the day of operation, 1 and 2 = the first 2 days following operation. Figures under (ff) represent subsequent determinations, at intervals after operation indicated by the figures in parentheses

of the series studied. The fall of serum albumin, in a larger series, is not as striking as is the fall of amino acids, because serum proteins are affected by disturbances of hydration, and because transfusions of whole blood or plasma are so often given in the postoperative period. This phenomenon had been reported earlier by Cuthbertson and Tompsett.[40] Such precipitate and sustained falls can not be referable to losses of blood or exudate during operation. It has been experimentally demonstrated that deficits of protein, resulting from far larger losses of blood than those sustained in most of these operations, can be replaced from the body reserves, in a short time. These are not reactions of surgical shock. They cannot be attributed merely to hemodilution, since globulin was only irregularly and temporarily involved.

More recently, Dr. Evelyn Man has shown that the serum lipids fall after operation. In this drop, all fractions of the lipids participate, without any significant disturbance of their relations to one another.

The explanation of these phenomena can be only conjectured, from such scattered observations. It is logical to regard them as the first evidences of the injury reaction, which later evidences itself in the "toxic destruction of protein." Although the experiments of Croft and Peters[24] can not be lightly dismissed, lack of a specific amino acid or other essential dietary constituent does not usually become apparent

so precipitately. Moreover, it should be not less, but more, evident in the chronically ill and wasted person, than in a previously healthy subject. The reaction appears to be characterized by impairment of the synthesis of protein, including serum albumin, and a tendency to route products of protein metabolism preferentially to urea and ammonia to be excreted in the urine. The self-terminative nature of the reaction has led to a revival of the old concept that the protein of the body can be divided into two categories, of which one is less essential and less solicitously protected than the other. It is, however, unnecessary to revive this hypothesis. It is a general characteristic of biological organisms to exert increasing resistance to forces that tend to distort their composition or their activities. Besides, the changes in the composition of the serum are not the results of protein losses; they precede these losses.

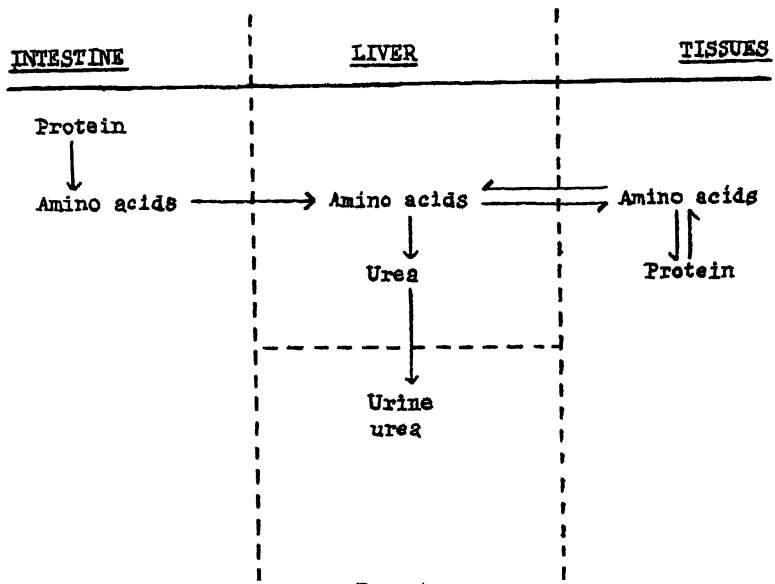


FIGURE 4.

In FIGURE 4, there is an oversimplified diagram of the course of protein metabolism. Protein is digested in the intestine to its component amino acids, which, after absorption, are carried to the liver and other tissues. In the liver, the amino acids may be broken down at once to urea and, as such, excreted, or they may be converted to other products, including new proteins. In the other tissues, they may enter new proteins or undergo transformation, while other amino acids, displaced from the tis-

sues, are returned to the liver. The special vehicles that constitute urea precursors are not separately indicated in the diagram, because they are measured with the amino acids by usual analytical procedures. In the phenomenon which we have been discussing, the path to synthesis in the tissues seems to be obstructed. The system appears to be short-circuited, in such a way that both dietary amino acids and amino acids derived from the tissues are shunted into the urine as urea. It is impossible to decide whether the primary derangement lies in acceleration of the destructive, or impairment of the synthetic, process. If the primary defect is accelerated metabolism, its origin should be sought in some disorder of hepatic function. The beneficial effects of methionine, then, might lie in its lipotropic action, or some associated property.

Since protein is an essential constituent of tissues which cannot be replaced by other foodstuffs, its loss is presumably deleterious and should be prevented. Obviously, this can be done effectively only by discovering and rectifying the disorder of metabolism responsible for the toxic destruction of protein. If, or as long as, it is impossible or impracticable to prevent or mitigate losses of protein after acute insults, a certain amount of wasting, with attendant delay of convalescence and rehabilitation, will have to be accepted as an inevitable consequence of such insults. This should, at least, be reduced to a minimum. In chronic debilitating conditions, synthetic powers of the body appear to be intact. The same is true, in acute pathologic states, after the catabolic phase is ended. Since the catabolic phase is self-terminative and of unpredictable duration, the advent of the anabolic phase should be anticipated by the administration of generous diets, before the catabolic phase has ended. The only contraindication to such a course would be the proof that such diets, during the catabolic phase, are positively injurious. This has not been established. An exception may have to be made of the few days immediately following a profound insult, such as a severe burn. These diets should contain, not only adequate amounts of protein, but also high calories.

The use of protein hydrolysates as substitutes for, or supplements to, diets has become the vogue. Although nitrogen equilibrium can be maintained, and the protein requirements of normal animals and human subjects can be satisfied, by injections of suitable hydrolysates, there is no evidence that such injections are superior to normal methods of eating. If hydrolysates are not injected at an extremely slow rate, they provoke vomiting. While they are being injected, patients will

seldom eat. If, therefore, they are used as dietary supplements, they must be given after the evening meal. This deprives the patient of much needed sleep. I would not belittle the value of these hydrolysates; but, in our experience, they are indicated only when there are positive contraindications to oral feeding, or when it is impossible to persuade a patient to take adequate amounts of nourishment by the normal route. For intravenous administration, mixtures of pure amino acids would have distinct advantages over less refined products, because they can be injected in high concentration, with great rapidity, and without provoking untoward reactions.[22] It is quite possible, also, that it will prove advantageous to vary the composition of such mixtures, to meet particular nutritive requirements. However, amino acid preparations suitable for such purposes are not, as yet, commercially available.

Intravenous amino acid mixtures or protein hydrolysates, as a sole source of nutrition after injury, are unsatisfactory from another point of view. There can be little doubt, from the experiments of Shaffer and Coleman[1] and others, that high calories, in the form of carbohydrate and fat, spare protein in this condition. Such calories can be given intravenously only in the form of glucose solution. The quantity of glucose that can be injected, in this manner, is distinctly limited.

It has been recommended that protein hydrolysates be administered by gavage. Tube feeding should be practised with caution, in patients who may vomit or who are unconscious, because of the danger of aspiration pneumonia. If it is practised, there is no reason to use protein hydrolysates. Milk fortified with milk powder is well tolerated, and cheaper in proportion to its food value. The necessary vitamins and other accessories may be added to the milk mixture.

Proper choice of foods and efficient nursing care are the chief essentials for success in feeding. In our experience, a resourceful team of nurses was able, by meticulous attention to details, to increase greatly the dietary intake of patients. If a sick or injured person will take fluids freely and has no gastrointestinal disorder, it is seldom impossible to introduce the necessary quantities of nourishment, by substituting highly nutrient for non-nutrient fluids. The ubiquitous pitcher of water may act as a deterrent to feeding. Milk, reinforced by milk powder and some sugar, slightly flavored with vanilla, coffee, or chocolate, is an excellent food. If a patient will not take fluids freely, the judicious administration of salt may provoke sufficient thirst to stimulate him to take, not only the desired amount of fluid, but also the proper amount of nourishment in these fluids.

Flavor and seasoning, especially salt, are conducive to both appetite and thirst. The impression that only unseasoned bland foods are suitable for sick people is inherited from the era when all food was considered harmful. Patients may be too weak and tired to contend with formal meals, though they would eat if the food were served in a less formidable manner, or if they were fed. Divided meals or supplementary feedings are useful, if intelligently planned. However, feedings are often so ineptly scheduled that the supplements only detract from the appetite for the next meal. Feedings just before retiring are customarily small; they can be enlarged to advantage. The fervor for more adequate feeding of patients must advance further than the order book. A patient does not profit from the diet ordered, but from the food he eats; the two become identical only when care is taken to translate the order into appropriate practical terms. Intravenous treatments must not be used as a means of evading this responsibility.

BIBLIOGRAPHY

1. **Shaffer, P. A., & W. Coleman**
 1909. Arch. Int. Med. **4**: 538.
2. **Kocher, R. A.**
 1914. D. Arch. klin. Med. **115**: 82.
3. **Coleman, W., & E. F. DuBois**
 1915. Arch. Int. Med. **15**: 887.
4. **Barr, D. P., & E. F. DuBois**
 1918. Arch. Int. Med. **21**: 627.
5. **Cecil, R. L., D. P. Barr, & E. F. DuBois**
 1922. Arch. Int. Med. **29**: 583.
6. **Coleman, W., D. P. Barr, & E. F. DuBois**
 1922. Arch. Int. Med. **29**: 567.
7. **Buell, M. V.**
 1919. J. Biol. Chem. **40**: 63.
8. **Endrus, G., & C. Neuhaus**
 1925. Z. ges. exp. Med. **47**: 585.
9. **Cuthbertson, D. P.**
 1932. Quart. J. Med. **1**: 233.
10. **Brunschwig, A., D. E. Clark, & N. Corbin**
 1942. Ann. Surg. **115**: 1091.
11. **Mulholland, J. H., F. Co Tui, A. M. Wright, & V. J. Vinci**
 1943. Ann. Surg. **117**: 512.
12. **Howard, J. E., W. Parson, K. E. Stein, H. Eisenberg, & V. Reidt**
 1944. Bull. Johns Hopkins Hosp. **75**: 156.
13. **Howard, J. E., J. Winternitz, W. Parson, R. S. Bigham, Jr., & H. Eisenberg**
 1944. Bull. Johns Hopkins Hosp. **75**: 209.
14. **Browne, J. S. L., V. Schenker, & J. A. F. Stevenson**
 1944. J. Clin. Invest. **23**: 932.
15. **Grossman, C. M., T. S. Sappington, B. A. Burrows, P. H. Lavietes, & J. P. Peters**
 1945. J. Clin. Invest. **24**: 523.

16. **Hirshfeld, J. W., W. E. Abbott, M. A. Philling, C. G. Heller, F. Meyer, H. H. Williams, A. J. Richards, & R. Obi**
1945. Arch. Surg. **50**: 194.
17. **Elman, R.**
1942. Ann. Surg. **112**: 594.
18. **Madden, S. C., & W. A. Clay**
1945. J. Exp. Med. **82**: 65.
19. **Cope, O., I. T. Nathanson, G. M. Rourke, & H. Wilson**
1943. Ann. Surg. **117**: 937.
20. **McCann, W. S.**
1922. Arch. Int. Med. **29**: 33.
21. **Rose, W. C.**
1938. Physiol. Rev. **18**: 109.
22. **Madden, S. C., J. R. Carter, A. A. Kattus, Jr., L. L. Miller, & G. H. Whipple**
1943. J. Exp. Med. **77**: 277.
23. **Kinsey, V. E., & W. M. Grant**
1944. Science **99**: 303.
24. **Croft, P. B., & R. A. Peters**
1945. Lancet **1**: 266.
25. **Waife, S. O., E. H. Malkiel, E. B. Man, & J. P. Peters**
To be published.
26. **Frisch, R. A., L. B. Mendel, & J. P. Peters**
1929. J. Biol. Chem. **84**: 167.
27. **Weech, A. A., M. Wollstein, & E. Goettsch**
1937. J. Clin. Invest. **16**: 719.
28. **Peters, J. P., & A. J. Eisenman**
1933. Am. J. Med. Sci. **186**: 808.
29. **Man, E. B.**
To be published.
30. **Cuthbertson, D. P., & S. L. Tompsett**
1935. Brit. J. Exp. Path. **16**: 471.

DISCUSSION OF THE PAPER

Dr. Sidney C. Werner (*Department of Medicine, College of Physicians and Surgeons, Columbia University, and Presbyterian Hospital, New York, N. Y.*):*

Dr. Peters has presented a very careful and well-informed discussion of the changes in nitrogen balance which follow trauma. He has outlined the catabolic and anti-anabolic theories, to explain the difficulty in establishing nitrogen balance which exists after injury, and shown the problems of analysis which exist.

The experience in our laboratory has been somewhat different from that of the New Haven group. During the past year and a half, a nitrogen balance study has been conducted on cases of peptic ulcer, subjected to partial gastrectomy, and on a miscellany of cases undergoing various surgical procedures. One case of severe fracture of the thigh treated in plaster was also observed. The operative patients were all given a mixture of pure amino acids, to provide 12 to 28 gms. of nitrogen per day, for the first three to five days after the procedure. This was supplemented by 75–100 gms. of glucose in saline. A diet was then given. The fracture patient received a casein hydrolysate and high protein milk, for the first three weeks after the trauma, up to 50 gms. of nitrogen intake per day.

There were twenty-one gastric cases treated by the above regimen, and six control cases treated identically, except that no amino acids were administered, and

* The work described in this paper was done under a contract, recommended by the Committee on Medical Research, between the Office of Scientific Research and Development, and Columbia University.

the usual postoperative diet was instituted after several days. Nine patients with other conditions were followed, and the control figures for nitrogen loss were obtained from the literature. Nitrogen loss was appreciably reduced in both groups. In the gastric group, 17 of the 21 treated cases lost less nitrogen, postoperatively, in the first five days, than any of the controls, and actual nitrogen balance was produced, in four gastrectomy patients, the fracture case, a case of appendectomy, and one of hernia repair. This is strong evidence against any anti-anabolic effect of injury. Another strong piece of evidence against this hypothesis is the continued high rate of excretion of nitrogen from tissue stores, when a high nitrogen intake is suddenly stopped. This should not occur, if the nitrogen wastage after trauma is purely anti-anabolic in origin.

There are objections to the catabolic theory, as well, first brought out by the early workers mentioned by Dr. Peters. Thus, a different interpretation from either has been necessary to explain the data obtained from these experiments, and this is being presented, elsewhere. In brief, it is thought that the mechanisms for anabolism and catabolism are not altered from the normal, except that both act at a higher rate of exchange. Thus, nitrogen equilibrium is attained only at a higher level of nitrogen intake than before the trauma. The height of this equilibrium point depends on the severity of the injury.

It should be mentioned that preceding weight loss is no sure index as to whether the body will react to trauma, or not, by increased nitrogen exchange. In the gastric cases mentioned here, the cases showing the most marked negative and positive balance showed somewhat similar weight losses before operation, varying from none to thirty pounds.

In conclusion, Dr. Peters' paper provides much food for thought and is an excellent example of the critical approach to the literature and data of the problem of nitrogen metabolism after injury. His conclusion, that every effort be made to prevent nitrogen wastage, is becoming the basis of present-day surgical therapy.

THE INTRAVENOUS USE OF PROTEIN AND PROTEIN HYDROLYSATES

By Robert Elman

Washington University, School of Medicine, St. Louis, Missouri

The intravenous route, as a method for the administration of fluids, was demonstrated as feasible, in 1658, by no less a person than Sir Christopher Wren, architect and astronomer. Within a decade after this experiment, transfusion of blood was carried out, from animal to animal and from animal to human, and aroused considerable excitement in scientific circles, in London as well as on the Continent. For example, it was thought that piety could be injected into an atheist, by using the blood of an Archbishop, or that an incompatible husband and wife could be made happy, by the transfusion of each other's blood. These unreasonable hopes, however, were not responsible for the failure of the method; the deleterious and even fatal reactions which followed finally led to its being outlawed. Even in 1818, when Blundell[2] first saved the lives of many women dying of postpartum hemorrhage, with transfusions of human blood, the method was still inapplicable, because of occasional untoward and fatal reactions, which we now know were due to the incompatibility of some of the four blood groups. But, even after these were discovered, in 1901, by Landsteiner,[11] transfusions did not achieve general use for three decades, partly because untoward reactions were still encountered. Untoward reactions were also responsible for the failure in the general use of intravenous injections of saline, which were first employed in 1831, and intravenous injections of glucose, which were first recommended, in the human, in 1896. Not until 1923, was the cause discovered for one of the most frequent of the harmful effects, *i.e.*, pyrogenic reactions. It was a woman scientist, Florence Seibert,[15] who demonstrated that a specific bacterium grew in distilled water and produced a stable substance which, on entering the circulation, could produce severe rigors and fever. It is interesting that Dr. Co Tui, whose paper follows mine, contributed much to this problem, and was greatly responsible for the fact that, at present, pyrogen-free water is universally available. It is no coincidence that the development of pyrogen-free water was followed by a tremendous increase in the use of the intravenous method for all sorts of solutions, including blood and plasma.

The intravenous route, as a method of administering various solutions, has developed very rapidly, in the past two decades. At the present time, the administration of amino acid mixtures, as protein hydrolysates, has been extensive, and many thousands of patients have undoubtedly received this form of therapy. Pure amino acid mixtures have also been used, but, thus far, the high cost has precluded their general use. The presence of unnatural forms of some of the amino acids has also been cited as a disadvantage in the use of such mixtures. Hydrolysates may be prepared in a variety of ways. Two such preparations are now available for general use. In the following paper, an enzymic hydrolysate* will be discussed, on the basis of an extensive clinical experience with this preparation. Before describing the beneficial clinical effects, the possible deleterious results will be discussed.

DELETERIOUS EFFECTS

As already mentioned, the clinical use of the intravenous route for the introduction of various solutions was long delayed, because of the untoward reactions following such injections. One of the commonest is the so-called pyrogenic reaction. It was really not until the last decade that pyrogen-free distilled water became universally available, and the importance of using pyrogen-free material and apparatus appreciated. If these dangers are eliminated, injurious effects which may follow intravenous injections are obviously due to the dissolved substance itself.

The dangers from the injection of protein hydrolysates were recognized early, and were so great that this method of treatment seemed doomed to failure. The first experiments[11] were carried out in 1889. Most of these harmful effects were described as peptone shock, because it was assumed that the peptones present in the hydrolysates were responsible. It is not clear, however, on the basis of present knowledge, whether they were due to the presence of peptones themselves, or to other materials, formed during the manufacture of these early preparations. In any case, by 1913, the successful intravenous use of a protein hydrolysate, in animals, was reported.[7] The elimination of deleterious effects was obviously due to better methods of preparation. At the present time, protein hydrolysates have been so extensively used in humans that considerable knowledge of the untoward effects, following intravenous injection, has been obtained. They will be discussed under the following subheadings:

* Amigen manufactured by Mead Johnson Company, Evansville, Indiana.

Amino Acids and Peptides

That amino acids, themselves, may be responsible for untoward reactions, was the assumption of Shohl and Blackfan,[16] who first injected a mixture of pure amino acids in the human. They found that pyrogenic reactions were just as likely to occur, with a protein hydrolysate, as with a mixture of crystalline amino acids of similar composition. However, these observers offered no data to show that pyrogens were not present in the distilled water used in making up the solutions. While it does not seem physiologically sound to incriminate amino acids and peptides, which are normally absorbed from the gastrointestinal tract during digestion, it was thought possible that the sudden entrance of relatively large amounts of these into the systemic circulation could provoke reactions, due, perhaps, to a suddenly increased dynamic action. At any rate, this idea has been subjected to considerable further study, by injecting various mixtures of pure amino acids into animals and in humans. In these studies, however, nausea and vomiting, rather than pyrogenic reactions, were observed. Nausea and vomiting are a reaction which seems peculiar to amino acids, inasmuch as solutions containing them in pure crystalline form, when injected at too rapid a rate, are capable of provoking such an effect.

Most of the animal studies on reactions following the intravenous injection of crystalline mixtures of essential amino acids have been made in the dog, by Madden and his co-workers,[12, 13] who found, however, that the best results were obtained when non-essential glycine was added, which improved the tolerance of the solution. On the other hand, two other non-essential amino acids were found to produce nausea and vomiting; these were glutamic acid and aspartic acid. A contrasting experience, with solutions of crystalline amino acids, was reported by Cox and Mueller,[7] who inferred from their observations that nausea and vomiting, in dogs, were much more likely due to the unnatural forms than to any particular amino acid.

In the human, the reported observations have been few. Those of Shohl and Blackfan[16] have already been mentioned. In another study,[1] in a surgical patient, a mixture of crystalline amino acids was given intravenously, for 30 days, without untoward effect, even when the rate of injection was rapid. The present author[5] has observed definite differences between the dog and the human, with the following mixture of pure crystalline amino acids:*

The mixture was made up, 75 grams per liter, *i.e.*, 7½ per cent, which

* Kindly supplied by Merck and Company, through the courtesy of Dr. D. F. Robertson, from a formula (Vu) made up by Dr. S. C. Madden.

TABLE 1

		Grams
dl	Threonine	7
dl	Valine	11
dl	Leucine	18
dl	Isoleucine	12
l	Lysine HCL	12
l	Tryptophane	4
dl	Phenylalanine	12
dl	Methionine	6
l	Histidine	4
l	Arginine	7
	Glycine	7
Total		100

contained 18 grams per cent of nitrogen. Glucose, 5 per cent, was added, before administration. When injected in dogs, considerable nausea and vomiting followed the injection of this mixture, as compared, for example, with a solution of hydrolyzed protein (Amigen). In humans, the results were just the reverse, *i.e.*, the pure amino acids were tolerated better, as shown by the fact that the solution could be injected far more rapidly without provoking nausea and vomiting, as compared with a solution of hydrolyzed protein (Amigen). Moreover, in three trials with a solution of pure glutamic acid alone (one liter contained 30 grams), no reaction was observed, following rapid intravenous injection, in convalescent patients.

Observations on the urine were made, in two surgical patients with esophageal obstruction due to carcinoma, following the intravenous injection of the above-mentioned amino acid mixture. Nitrogen balance was not greatly different from that obtained with the same amount of nitrogen as hydrolyzed protein. But the excretion of amino acids was quite different, as shown in TABLES 2 and 3. Roughly 20 per cent

TABLE 2

CARCINOMA OF THE ESOPHAGUS

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
N intake	None			Mixture Vu								None				Amigen				None	
Amino N, % of urine N	3.5	2.2	2.8	7.6	11	8.3	18	16	18	20	17	8.1	2 5	2.6	2.5	4.2	4 9	4.4	4.9	2.5	2.8

TABLE 3

CARCINOMA OF THE ESOPHAGUS

Day	Nitrogen intake	Urinary N	
		% NH_3 & urea	% Amino N
1	0	84	1.6
2	0	—	—
3	0	—	—
4	Vu	66	16
5	Vu	—	—
6	Vu	62	20
7	Vu	63	10
8	Vu	64	22
9	Vu	63	22
10	Vu	62	25
11	Amigen	79	4.5
12	Amigen	81	2.4
13	Amigen	80	6.4
14	Amigen	83	1.6
15	Amigen	79	2.5

of the injected nitrogen of crystalline amino acids appeared in the urine as amino acids (presumably unnatural forms), as compared with a few per cent, in the case of the hydrolyzed protein.

Other Substances

Undoubtedly present in hydrolysates, but not in pure mixtures of amino acids, are other substances which may be responsible for some of the harmful reactions. This was undoubtedly true of many of the earlier preparations of protein hydrolysates, which seemed to provoke a higher incidence of untoward effects, as compared with present preparations, which are, presumably, purer. A specific study was made by Hopps and Campbell,[8] who showed, by *in vitro* tests with guinea pig muscle, that enzymic, as well as acid, hydrolysates of casein contain substances assumed to be either like histamine, peptones, or tyramine. Their presence would seem to explain such clinical phenomena as flushing of the skin, various vasomotor and other signs and symptoms, whenever the injection rate is too rapid. Further observations are needed, in order to study the effect, and eliminate as much as possible of the substances of this nature.

Products of Bacterial Contamination

Bacteria and/or the products of bacterial growth may readily be responsible for severe pyrogenic and other reactions, which may even be

fatal. Such substances may readily form, particularly because solutions of hydrolyzed protein are excellent culture mediums. It is probable that contamination may result in such rapid bacterial growth, within a relatively few hours after the sterile solution is opened to the air, that reactions may occur. Suppose that a few bacteria gain entrance and begin to multiply immediately. Eventually, of course, the solution becomes turbid, and the evidences of contamination are readily apparent upon inspection. There is, however, a period before the solution becomes turbid, in which bacterial growth has been considerable. The injection of such a clear solution is, obviously, fraught with danger; yet, it may readily be avoided by the simple precaution of injecting the clear solution, as soon as the sealed container is opened. One liter of such a solution can be given to an average sized adult within two or three hours. It is unlikely that sufficient bacterial growth could take place, within this period, to produce dangerous reaction. On the other hand, if such a flask is opened and 6 to 8 or more hours allowed to elapse before the injection is started, such growth could occur. It may be that a few untoward and, perhaps, fatal reactions may have been caused by such a circumstance. The obvious lesson is the avoidance of contamination by: (a) the use of closed systems, similar to those recommended for all intravenous injections; (b) starting the injection immediately after the container is opened, and completing it within a period of 3 or 4 hours at the most.

Allergens

Solutions of hydrolyzed protein may contain substances to which a particular patient is sensitive. This, undoubtedly, occurs, even though complete and thorough tests show that no anaphylactic substances are present in the product. Allergens comprise a variety of things and are the obvious cause of urticaria, following an intravenous injection. Other manifestations of allergy are severe itching of the skin, patchy edema often called angioneurotic edema, and skin rashes. Each of these has been observed occasionally. The author once saw a pronounced edema of the mesentery, at operation, in a patient who, within the previous 24 hours, had received an injection of hydrolyzed protein without apparent reaction at the time. Serious allergic reactions of this sort have been rare, although it is possible that a fatal result may occur. The incidence is not known, but must, naturally, be compared with the fatal reactions which have been observed following blood and plasma transfusions, which, even in recent years, have been reported as occurring about once in every one to three thousand transfusions.

Chemical Incidence of Reactions

Actual experiences with the rate of reactions, following the intravenous injection of solutions of hydrolyzed protein, have been published, but most other reports deal with single reactions or the analysis of a relatively small number of cases. The author[5] has carefully studied each consecutive injection, during a period of 9 months, the results of which are presented in TABLE 4.

TABLE 4

Total number of patients receiving Amigen	352
Average injection per patient (each 1 liter)	8
Largest number of injections in one patient	53
Total number of injections (each 1 liter)	2,729
Total number of untoward reactions	22 (0.8%)
Nature of Reactions	
Pyrogenic (chill, with or without fever, or fever without chill)	11 (50%)
Allergic (itching, urticaria, rash, edema)	7 (31%)
Miscellaneous	4 (19%)

The four miscellaneous reactions were described as disturbing, subjective sensations, such as tingling and suffocating,[1] vomiting,[1] faintness and perspiration,[1] headache and dizziness,[1] all of which seemed severe enough to discontinue the injection. Of the other 18 reactions, 12 were observed in patients who received Amigen, previously and subsequently, without incident. Of the total of 22 reactions, in only 10 was the injection a first one and did the reaction constitute a contraindication. In many of these, no reactions might have been observed, had the injection been repeated. In two of these, reactions occurred with subsequent injections of other fluids. There were no deaths attributable to the injections. The speed of the injection was somewhat slower than that of other intravenous infusions. In general, 2 hours was the average duration for one liter. With faster rates, nausea and vomiting were encountered occasionally, and this was not recorded as a reaction, unless they persisted and the infusion was discontinued. In all others, the symptoms disappeared as the flow was slowed down.

Other Deleterious Effects

Regarding phlebitis and thrombosis, the influence of hypertonic solutions is well known. The influence of pH seems to be particularly important, with solutions of amino acids, because of their great buffering properties. Experimental study has shown that phlebitis is more likely,

following injections of solutions of amino acids, with an acid reaction.[9] Another possible harmful effect of injecting a solution of amino acids, at a low pH, is the fact that they are capable of significantly lowering the carbon dioxide-combining power of the blood. For example, solutions of hydrolyzed protein, at a pH of 5.5, may significantly reduce the CO_2 combining power of the blood, whereas, when neutralized to a pH of 6.5, it has but slight effect on the CO_2 combining power and none on the pH of the blood, as shown by actual measurements in two patients.[8] After the injection, during 8 hours of 3 liters of a 5 per cent solution, there was a drop of but 6 and 8 volumes per cent, respectively. This is another reason why amino acid solutions should not be given at a pH which is very far from the neutral point.

CLINICAL FINDINGS

The advisability of using a solution of hydrolyzed protein would, obviously, depend upon whether it is of value in the treatment of disease, once it has been shown to be safe. If food is of value, whether it is injected intravenously or taken by mouth, it would seem that the only requirement in using a solution intravenously would be: Is the material injected used as a food? However, the assumption that food is beneficial, in the treatment of disease, does not have many adherents. In other words, starvation during disease is deeply ingrained in the tradition of the healing art; not only is it held to be inevitable, but even necessary, at least during certain periods in a patient's illness. Much of this belief follows a blind reliance on the patient's appetite. Most physicians apparently feel that if the patient does not want to eat, food is bad for him; that, as soon as he is well enough to eat, he will start eating.

Aside from caloric restriction in certain obesity diets, there is no evidence that starvation has any beneficial effects. Indeed, the overwhelming evidence points to a contrary conclusion. As a result of the traditional *laissez-faire* policy, in the nutritional care of the sick, many medical and surgical wards are filled with patients, just as many unhappy countries of Europe are filled with people, who are showing numerous harmful manifestations of starvation.

In presenting our clinical data, therefore, I would like to assume that starvation is injurious, that food is something which should be considered as beneficial, and that the question of introducing intravenous protein nourishment, as amino acid mixtures, reduces itself to the question of whether it is really a food or not. Against this assump-

tion, must be mentioned the evidence, described by Dr. Peters, that protein food cannot be assimilated, even when ingested by mouth, during certain catabolic periods, following injury and infection in a well-nourished individual. Under such a condition, there would seem to be little justification for administering protein, regardless of the channel or method used. In contrast to such findings, we have observed but little evidence, in our patients, that this was true after serious abdominal operations. Moreover, the data presented by Brunschwig, by Lund, by Gardner, and other surgeons all indicate[3] that considerable retention of nitrogen does occur, during the postoperative period. As suggested by Dr. Peters, this difference must, obviously, be due to differences in the kinds and conditions of patients studied. Entirely apart from the previous nutritional condition of the patient, one might cite at least two differences between a patient who suffers an accidental injury and one who undergoes an operation of choice. In the first instance, excitement is inevitable, pain unavoidable, and unless an operation is carried out, no anesthesia is used. In the second instance, excitement is usually absent, no pain is inflicted, and the anesthesia itself exerts definite influences. It is probable that the influence of adrenal cortical stimulation, the so-called "alarm reaction" following injury, is much greater, following an accidental injury, than that following operation.

Regardless of these considerations, however, it is doubtful in my mind whether the existence of such a catabolic phase should justify a planned policy of starvation. For one thing, it would be difficult to tell when this phase was over. It would seem to be better to feed the patient moderately, at least, so that the nutriment may be utilized as soon as the body is able to do so. In this way, a policy of prevention is adopted, from the beginning, which will certainly tend to avoid many of the severe instances of malnutrition which are so often seen after injury or operation. In the meantime, study should be continued of this catabolic mechanism, so as to minimize or remove it, as much as possible. However, these considerations are not peculiar to intravenous feeding; they apply, equally, to oral alimentation.

In presenting our findings, the importance of an obvious, but often forgotten, policy should be emphasized: to use the parenteral route only when the oral route cannot be used, and to use the parenteral route for as short a period as possible. Indeed, I would say that the most successful program of parenteral alimentation is the one which ends the soonest; in other words, which restores the patient to normal condition as soon as possible. This, indeed, has been our experience with the use

of a more complete intravenous diet. I would advocate the use of more complete solutions for intravenous feedings, but less of them. Lack of space prevents discussion of the many indications for the parenteral use of amino acid mixtures. In my practice, I use hydrolyzed protein in any patient who cannot eat or drink and would ordinarily receive the usual injections of glucose and saline. In presenting our findings, however, the observations will be confined to patients after operation only. In these cases, the patient cannot take food by mouth, for a few days, largely because of nausea and vomiting, and often because we wish to keep the gastrointestinal tract at rest, until it is able safely to absorb and digest fluids and food taken by mouth.

In our experience, we have observed only four possible contraindications to the use of hydrolyzed protein intravenously (but are not sure that the third and fourth are justified): (1) a solution which has been open for several hours or is not crystal clear; (2) an untoward reaction of the allergic type; (3) severe hepatitis; (4) postoperative anuria.

The beneficial effects of including hydrolyzed protein in the solutions given parenterally, after operation, were evident, both objectively and subjectively, as compared with patients receiving the usual postoperative therapy. The detailed observations have been described in a previous report.[6] While part of the improvement in surgical convalescence may have been due to psychotherapy and early movement, the improved nutritional status undoubtedly played an important part. Patients who received hydrolyzed protein, in their postoperative parenteral injections, were stronger and showed less anorexia than those who received glucose solutions, alone. This has also been the experience of Davis,[4] who observed patients receiving both forms of therapy.

Objective measurements showed that the weight loss was less than the controls (one pound as compared with nine pounds). Determinations of serum albumin concentration showed a slight increase, as compared with a slight decrease in the controls. Observations were also made on nitrogen excretion, which showed a negative balance of but one to four grams a day, in patients receiving hydrolyzed protein, as compared with those given glucose, only. TABLE 5 shows additional

TABLE 5

NITROGEN BALANCE DURING 76 POSTOPERATIVE DAYS

All serious abdominal operations

Total Nitrogen output urine	1142 grams
Total Nitrogen intake*	861 grams
Average daily Nitrogen loss	3 7 grams

*As intravenous Amigen *only*, exclusive of plasma and whole blood transfusions. Patients took nothing by mouth; caloric intake was intravenous glucose 400.

data, *i.e.*, in an unselected group of patients receiving hydrolyzed protein, the postoperative negative nitrogen balance, in 76 postoperative days in which collections were complete, averaged 3.7 grams. From previous observations, in similar cases receiving glucose, only, the negative balance would have been 10 to 20 grams.

In TABLE 6, are listed the urinary nitrogen determinations, in a well-nourished patient after a resection of the sigmoid, in which positive nitrogen balance was achieved.

TABLE 6

POSITIVE NITROGEN BALANCE AFTER RESECTION OF SIGMOID

Caloric intake, 1000 calories. Patient well nourished. No postoperative complications.

P.O. Day	Nitrogen intake		Nitrogen output, urine	Creatinine	Balance
	Food	I.V. Amigen			
1	0	6 0	6 7	1.75	−1 7
2	0	9 0	—	—	—*
3	0	12 0	9 2	1.97	+2.8
4	0	12 0	11.2		+1.8
5	0	12 0	12 8	1 09	−1 2

*Incomplete specimen

SUMMARY

The probable causes of deleterious effects, following the intravenous injection of amino acid mixtures, including hydrolyzed protein, have been discussed, and the clinical reactions, which totaled 0.8 per cent in a series of 2,729 consecutive injections, analyzed.

The beneficial clinical effects of adding hydrolyzed protein to the usual postoperative injection of glucose are described, in comparison with controls. These effects include: bedside improvement in surgical convalescence; elimination of, or decrease in, the negative nitrogen balance; less loss of body weight; and an increase, rather than a decrease, in serum albumin.

BIBLIOGRAPHY

1. **Bassett, S. H., R. R. Woods, F. W. Shull, & S. C. Madden**
1944. Parenterally administered amino acids as a source of protein in man. New England J. Med. **230**: 106.

2. **Blundell, J.**
1939. An exhibition of books on the growth of our knowledge of blood transfusions. Bull. N. Y. Acad. Med. **15**: 622. (Quoted by **G. L. Annan.**)

3. **Cox, W. M., Jr., & A. J. Mueller**
1944. Intravenously administered crystalline amino acids in causation of vomiting in dogs. Fed. Proc. **3**: 56.

4. **Davis, H. H.**
1945. The routine use of protein digest intravenously, following major surgical operations. Surg., Gynec., & Obst. **31**: 81.

5. **Elman, R.**
Parenteral Alimentation. P. B. Hoeber. New York.

6. **Elman, R.**
1945. A design for surgical convalescence. Ann. Surg. **122**: 716.

7. **Henriques, V., & A. C. Anderson**
1913. Über parentale Ernahrung durch intravenöse Injektion. Ztschr. f. Physiol. Chem. **88**: 357.

8. **Hopps, H. C., & J. A. Campbell**
1943. Immunologic and toxic properties of casein digest as prepared for parenteral administration. J. Lab. & Clin. Med. **28**: 1203.

9. **Horwitz, A., L. A. Sachar, & R. Elman**
1943. An experimental study of phlebitis following venoclysis with glucose and amino acid solutions. J. Lab. & Clin. Med. **28**: 842.

10. **Kilduffe, R. A., & M. DeBakey**
1942. The Blood Bank. The Technique and Therapeutics of Transfusions. C. V. Mosby & Co. St. Louis, Mo.

11. **Landsteiner, K.**
1901. Über Agglutinationserscheinungen normalen menschlichen Blutes. Wien. Klin. Woch. **14**: 1132.

12. **Madden, S. C., *et al.***
1944. Amino acid mixtures effective parenterally for long continued plasma protein production, casein digests compared. J. Exp. Med. **79**: 607.

13. **Madden, S. C., *et al.***
1945. Plasma protein production influenced by amino acid mixtures and lack of essential amino acids. J. Exp. Med. **82**: 77.

14. **Neumeister, R.**
1889. Zur Frage nach dem Schicksal der Eiweissnahrung im Organismus. Sitzungs-Berichte d. Physik. Med. Gesell. Würzburg. **1889**: 64.

15. **Seibert, F. B.**
1923. Fever-producing substance found in some distilled waters. Am. J. Phys. **67**: 90.

16. **Shohl, A. T., & K. D. Blackfan**
1940. The intravenous administration of crystalline amino acids to infants. J. Nutrition **20**: 305.

DISCUSSION OF THE PAPER

Dr. Everett I. Evans (*Medical College of Virginia, Richmond, Va.*):

Dr. Elman is to be congratulated for his excellent studies, and his perseverance in elaborating and proving the clinical worth of amino acid preparations for intravenous infusion. The occurrence of unsuspected low tissue and plasma reserves of protein, in surgical patients, is far more widespread than, I believe, we have hitherto thought. In fact, in studies made by Dr. J. C. Forbes and myself, on a large group of ward patients being brought into the hospital for elective surgery, we found, by plasma volume studies, a deficit in total circulating albumin in about 50% of these patients. Further, in a larger group (80%), there was found a more or less serious deficit in total circulating hemoglobin. The true status of these findings was only appreciated when it was found that the majority of the patients were thought to have normal values for circulating albumin and hemoglobin, because, by routine laboratory tests, the majority of the values were in the normal range; the actual values for albumin and hemoglobin were low, because of a moderate, to severe, deficit in total circulating plasma volume.

These findings make it certain that far more patients coming to surgery need

preoperative management, with amino acid (and other) solutions, than was evident by the ordinary laboratory investigations. Therefore, we believe there is a wide use for amino acid preparations in surgical patients.

Some of the shortcomings of amino acid therapy should be mentioned. One matter that disturbs some of the patients is that intravenous therapy requires a moderate period of immobilization of the arm. A much smaller group of patients complains of nausea, during the infusion of amino acid preparations. We have found it highly important not to mention to any patient who is to receive amino acid infusions that he might develop signs and symptoms of nausea and vomiting. Indeed, if one does not alarm the patient about these problems, we have found that very few patients, now, have any difficulty with nausea and vomiting. Further, we are finding that one can give amino acid preparations much more rapidly, in most instances, than we had thought, formerly.

Some mention should be made of the possible toxicity of amino acid preparations. One hears, from some persons, of the probability of five to six deaths, after amino acid had been given intravenously. So far as I know, there is only one recorded death in the literature; to this, we must add one other. In fact, our single death came with the first infusion of Amigen that we gave. The patient was quite seriously ill, was severely jaundiced, and died after receiving only five to ten cc. of the preparation. Autopsy did not reveal an evident cause of death. However, the symptoms immediately preceding death caused us to believe that the disturbance was allergic in nature.

Although we are convinced, in our clinic, that there is a real need for amino acid therapy in many surgical patients, we would be remiss in not mentioning the critical need, in practically all our patients, for whole blood. Indeed, anemia of the secondary type is one of the most serious problems. Unfortunately, of course, amino acid preparations will not take care of this anemia, although there can be little doubt that they aid in the later elaboration of hemoglobin.

In the postoperative period, we feel there will be probably less need for amino acid therapy if every effort is made to mobilize the patient, early after operation. If one can get patients out of bed, by the third or fourth day, and supply adequate, appetizing diets, there will be little need for amino acid therapy, in most patients. Our goal should be to get patients early out of bed and early out of the hospital. I might add, for what it is worth, that I am firmly convinced that there is need for a revolution in the preparation and serving of food to hospital patients. It might be well for hospitals to employ more good chefs and cooks and fewer dieticians.

I am sure Dr. Elman will agree that, if properly selected and prepared food can be taken by mouth, there will be little need for amino acid therapy. We have had exceedingly gratifying results, in burned patients, by supplying a diet, composed of meat protein, at a level of 150 grams a day. This diet seems to have done all that any amino acid preparation could do.

Finally, it is my impression that care should be taken, by biochemists, physiologists, and clinicians, that a new fad is not made of amino acid therapy, as has been done with vitamin therapy. As Dr. Elman has shown, many patients have a real need for this new therapy. It should not be given out of deference to mere publicity.

CLINICAL EXPERIENCES WITH ORAL USE OF PROTEIN HYDROLYSATES*

By Co Tui

Laboratory of Experimental Surgery, Department of Surgery, New York University College of Medicine; New York University Division of Surgery, Bellevue Hospital, New York, N. Y.

The oral use of protein hydrolysates, clinically, is based mainly upon the following considerations:

1. There is increased excretion of urinary nitrogen, in disease and following injury.

2. Proteins may be lost in exudates from open lesions and in any discharge.

3. Pathological processes are usually accompanied by anorexia or other gastrointestinal disturbances, which reduce food intake and absorption.

4. The hospital dietary is such that: (a) as a result of economy, the diet may be sub-standard; (b) as a result of the same administrative attitude, the serving of a high protein diet is the exception rather than the rule, for the average patient; (c) the food served is usually inadequate for the replenishment of proteins lost; (d) intravenous alimentation is a mitigating, rather than a fully adequate, measure for the provision of the body with proteins.

Factor 3 is self-explanatory, but the other factors need some further explanation.

Increased Urinary Excretion in Disease and Injury

This phenomenon has been known for some time, but attention has been redirected to it by the recent work of Cuthbertson.[1] Whereas the nitrogen excretion in the urine of the average individual on a mixed diet approximates 13.5 grams daily,- in the diseased patient it may rise several-fold. This phenomenon has found recent confirmation, in this country, in the work of Albright,[1] Browne,[1] Lund *et al.*,[3] Howard,[6-8]

* The work described in this paper was done under a contract, recommended by the Committee on Medical Research, between the Office of Scientific Research and Development, and New York University.

This paper represents the work, not only of the author, but also of a group of collaborators, among them: Doctors Arthur M. Wright, John H. Mulholland, Irving Barcham, George R. Gerst, N. Habaluyas Kuo, Gerald T. McCarthy; the nursing staff: Mary Coughlin Shirley Johansen, Helen Koestner, Mary Fluellen, Ruth Hullopeter; the chemical staff: Mrs. Lilly Schmidt, Mrs. Dorit Essler, Frieda Wolf, Sally Sells, Louis Sutherland, Arthur Wallen, and Rosa Kommer.

and Peters.[9] It has been called "toxic loss of nitrogen," and the period during which this loss is most acute, the "catabolic period." It has been ingeniously related, by Albright and Browne, to the "alarm reaction"[3, 4] described by Selye, and one of the cortical hormones of the adrenals has been thought to play a part. This loss of nitrogen was, at first, thought to be incorrigible: that is, regardless of how much protein is fed to the patient during this period, he cannot utilize it and is, therefore, doomed to protein wastage. At first, this catabolic period was thought to last several weeks, but it has now been reduced, by the same protagonists of the theory, to a few days, thus removing to the academic field this major obstacle to full protein replacement during the convalescent period.

That this incorrectable period does not occur with any consistency, at least not in surgical cases, may be seen from TABLE 1, which sum-

TABLE 1

CRITICAL RANGES OF NITROGEN INTAKE IN SURGICAL CONDITIONS

Surgical condition	N. Intake gm KBW	Cases studied
Gastrectomy	.27 –.42	25
Cholecystectomy	224–.339	9
Herniotomy	.147–.182	18
Appendectomy	.184–.359	7

marizes the study of nitrogen balance, for the first seven to twelve post-operative days, in four surgical conditions; namely, gastrectomy, cholecystectomy, herniotomy, and appendectomy. It was found that, while herniotomy cases achieved positive nitrogen balance, when the level of protein intake lies between .147 and .182 grams nitrogen per kilogram body weight, appendectomy cases require .184 to .359 gm. N KBW; cholecystectomy cases, from .224 to .339 gm. N/KBW; and gastrectomy cases, .27 to .42 gm. N/KBW. This protein intake was in the form of protein hydrolysates, given with dextri-maltose, so that each patient received 40 calories per kilogram per day.

This finding suggests that there may be a critical zone of protein intake, specific for each disease condition, below which zone patients go into negative balance, and above which, into positive balance; and that workers who used natural food in their studies could not supply nitrogen at a high enough level to achieve nitrogen equilibrium. In the gastric cases, in whom the study included body weight determinations and bedside ergography, the nitrogen gain was accompanied by an early return of strength, by gain in weight, and the ability to get out

of bed sooner, while in those registering nitrogen losses, there were losses in body weight, weakness, and prolonged convalescence. The latter suggests that the weakness and debility, following disease and operation, may not all be due to disease itself and may be partly or mainly of nutritional origin.

Protein Loss in Body Fluids

The loss of all body discharges is accompanied by loss of protein or of products of protein breakdown. Thus, there is nitrogen loss in the urine, in the feces, and in the sweat. The discharge from fistula may carry out of the body unabsorbed food matter, and the loss of saliva or of the cerebro-spinal fluid would drain from the body the amount of protein present in these two fluids. The protein loss from denuded surfaces in burns and from large discharging cavities, such as in lung abscesses and empyemata, may lose for the body as much as 60 grams of protein, in 24 hours.[10] This amount is well over one-fourth of the total circulating body proteins; it corresponds to over four plasma transfusions, and may account for the wasting, cachexia, and high mortality, in these conditions.

The Hospital Dietary

In most hospitals, the diets are an automatic matter, left in the hands of the dietician and given the minimum of medical supervision. In fact, the medical staff, having had only a rudimentary nutritional training, are glad to leave nutritional matters in the dietician's hands. This lack of understanding of the importance of proteins to bodily well-being may be laid to the teachings of Chittenden,[11, 12] whose work may be epitomized as engendering a philosophy of scarcity, rather than one of abundance. The clinician exposed in his student days to such teaching carries away the impression that, since the protein needs of the body are so easily met, it is a matter that requires little attention from him.

The Sherman diets,[13] recently re-approved by the National Research Council,[14] were formulated for the well individual. How they became transferred, without any question, to the hospital, as diets for the convalescent, has never been explained. The "basic" or "regular diet" contains 70 grams of protein (11.2 grams nitrogen), one gram per kilogram body weight. This diet, supposedly minimum, more frequently than not becomes maximum for the patient. The high protein diet contains 120 grams of protein (19.2 grams nitrogen) in civilian hos-

pitals, and 130 grams (20.8 grams nitrogen) in military hospitals; but, owing to the necessity of economy, it is almost impossible for the average hospital physician to prescribe a high protein diet for the average case.

In view of the loss of nitrogen in disease and injury, it is understandable that the "regular" diet is scarcely ever adequate, while the high protein diet may, in some severe cases, also be insufficient. Thus, a person losing 10 grams of urinary nitrogen, to take a low amount, would certainly need more than 11 grams of nitrogen in the diet to replace this loss, if one remembers that it takes more than 1 gram of protein for the replenishment of each gram lost. In a person who loses 18 grams in the urine and 6 in a discharge, even the high protein diet becomes markedly inadequate.

To make matters worse, many hospitals, in an attempt at effecting economy, reduce the regular diet further, as has been shown in a previous report.[15]

While it would be helpful, for hospitals, to increase the protein content of their regular diets to 100 grams and that of their high protein diets to 150 grams, yet the ceiling of protein intake, in natural food, is so low and so limited by the factor of satiety, that this measure in itself would not correct the tendency to malnutrition in the average hospital patient. The protein hydrolysates, because of the large amount which can be administered, admirably bridge the gap between the nitrogen intake and output, in the convalescent patient.

The Inadequacy of Intravenous Alimentation

Intravenous alimentation with preparations of protein hydrolysates is an outstanding contribution, in the field of convalescent nutrition. The credit for this must be ascribed to Elman,[16] for his pioneer work, and to Cox and his associates,[17] for making the products available for widespread use. However, intravenous alimentation was never intended to be any more than a mitigating measure during the period of large protein losses, when the patients are unable to tolerate the ingestion of food.

With the present low concentration of nitrogen and glucose in intravenous preparations, it is almost impossible to achieve nitrogen and caloric requirements, solely through intravenous feeding. A little calculation will make this evident. The practical physiological limit for volume intake intravenously is, at the maximum, 3000 cc. daily. With a solution containing 5% each of protein hydrolysates and dextrose,

the patient receives only 150 grams each of the hydrolysate and of dextrose. This furnishes the patient with 18 grams of nitrogen and less than 1200 calories daily. While the protein supply may be adequate in a mild case, the caloric supply is only two-thirds of the minimum requirement. It may be that a 7.5% solution of such a hydrolysate, in a 10% glucose solution, will provide a safe nitrogen and caloric margin, but then the solution may become too thick to administer conveniently; the time of administration becomes correspondingly prolonged, adding to the inconvenience of the patient and to the danger of thrombosis at the site of injection, a danger which is already enhanced by the greatly increased hypertonicity of the solution. Elman[16] has advanced the opinion that, in previously well-nourished patients, nitrogen balance can be attained without attaining caloric satisfaction, the patient burning up his own fat stores for the latter purpose. This, however, is no solution to the nutritional emergencies which arise, when a previously depleted patient undergoes an operation or falls a prey to disease.

In view of the inadequacy of intravenous alimentation and the dangers and inconvenience attending it, it should not be looked upon as a method of choice, but should be reserved only for that period in which the patient is unable to tolerate oral alimentation, and must be discontinued, in the latter's favor, as soon as possible.

Hyperalimentation in Peptic Ulcers

The treatment of 30 cases of intractable peptic ulcers, using the hyperalimentation method employing protein hydrolysates and dextrimaltose, has been reported in a previous paper.[18] All but three were hospital cases. Since then, 63 additional cases have been accumulated, this time treated ambulatorily and with four different brands of protein hydrolysates. As in the previous series, all 63 were intractable cases.

It is relevant, at this point, to explain why only intractable cases were chosen. Many peptic ulcers have a tendency to spontaneous remission, so that, in an indiscriminate series, it would be difficult to tell whether the improvement brought about was due to this tendency to remission or to the effect of the therapeutic regimen. With intractable cases, however, this doubt is largely removed. These 63 patients had suffered from the condition for periods varying from 3 to 19 years. All of them had had several x-ray examinations and were usually so medically sophisticated as to be able to cooperate intelligently in carrying out the treatment under supervision. There was never any doubt

in their minds which measures brought them relief. Most of them, disillusioned with what the medical profession could offer them in the way of relief, had despaired of ever being helped.

These patients were given the hydrolysates to take, for the first 24 to 48 hours, and instructed to report by telephone. If they did well on the first type of hydrolysate, for the first 48 hours, they were given enough of the same material to last them two weeks, but told to report as soon as they had any trouble, or, otherwise, once a week.

All natural food was withheld, and resumed if the patient had no pain for two weeks. Smoking, and drinking alcoholic beverages were prohibited. Patients were allowed to go to work, unless the work imposed too much of a strain on them, physically or mentally.

The hydrolysates were given in amounts corresponding to 0.6 grams of nitrogen per kilogram body weight. The caloric equivalent of the hydrolysate was calculated, and the balance of the caloric intake was made up by salt-free dextri-maltose, so that the patient was taking from 40 to 50 calories per kilo.

Both powders were dissolved in from 1 to 1.5 liters of hot water, and stirred until as much as possible would go into solution. The mixture was kept in the refrigerator to prevent souring. It was divided into 8 feedings, given every two hours during the waking time. For those patients having night pains, an additional feeding was given at night. Occasionally, a patient would have distress before the two-hour period ended. In that case, the mixture was divided into twelve feedings, given every one and a half hours. Some patients prefer to dissolve the hydrolysate and the dextri-maltose separately, taking the hydrolysate first and then following up with the dextri-maltose as a "chaser."

In this series of patients, we have found kaolin, in the form of kao-magma or kaopectate, more efficacious, in the treatment of diarrhea, than amphojel, which had been used in the series previously reported. No vitamins were given until the second week of treatment, when a full complement was administered.

After two pain-free weeks, the patient was put on a bland diet, with the usual three meals, but so timed that they took the place of three of the hydrolysate feedings. The patient thus kept on taking five hydrolysate feedings daily, without dextri-maltose, for as long as he could tolerate hydrolysates. When tired of hydrolysates, he was asked to substitute dried milk, two or three tablespoons in hot milk, every two hours, except during meal time.

The data of this series of 63 cases of peptic ulcer, treated ambulatorily, may be briefly summarized as follows.*

Sex: 61 male, 2 female.

Age: 11 to 58 years.

Type: 2 gastric, 60 duodenal, 1 recurrent (postoperative gastrectomy).

History: Night pains in 18, melena in 5, hematemesis in 2, perforation in 3.

Results: Excellent in 47, good in 12, failure in 4 (1 hemorrhaging under treatment and 3 unrelieved, including the recurrent).

Excellent: Pain stopped immediately, in 6; in 24 hours, in 35; in 48 hours, in 6. Night pain disappeared, after the first or second days, in all; nervousness disappeared, in from 4 days to 1 week.

Good: Pain stopped, in from 5 days to 10 days. There was gain in strength and morale, and a greater inclination to work.

In this series, there were three failures, one of them a gastro-jejunal ulcer.

Again, it must be reiterated that this regimen is only a means of bringing a symptomatic peptic ulcer condition to a prompt remission and roentgenologic healing. No claim is being made for a permanent cure, nor can it be said to touch the psychosomatic picture of the disease, unless the restoration of cheerfulness and increased morale may be said to bring about an interruption of a vicious circle, this time in a reverse way, a somato-psychic beneficial circle.

In a previous report, two reasons were adduced for the prompt response to the hyperalimentation regimen:

1. The neutralization of the gastric juice;

2. The feeding of a nutritious mixture which furnishes the body with repair material.

In the light of our experience, *i.e.*, of better results from more complete hydrolysates than from less complete ones and dried milk, a third reason may be suggested: that the relief of digestive effort on the part of the gastro-duodenal tract was also conducive to healing.

* The author gladly acknowledges his debt of gratitude to Mead Johnson and Company, for supplying Amigen for the treatment of ten of the cases in this series; E. R. Squibb and Sons, for supplying their protein hydrolysate for the treatment of thirty cases; Sheffield Farms Co., Inc., for supplying their product, Edamin, for twelve cases; National Drug Company, for supplying their product, Aminonat, for eight cases; and Frederick Stearns and Company, for supplying their product for three cases.

Advanced Pulmonary Tuberculosis

(Under the supervision of Doctors J. Burns Amberson and Julia M. Jones, Chest Service, Bellevue Hospital.) Two cases of advanced tuberculosis, who were put on the hyperalimentation regimen, were reported in a previous paper.[19] Both improved in weight and in physical well-being, but fifty-seven days after the start of the regimen, there was still no roentgenologic improvement of the pulmonary lesions. It may now be reported that monthly roentgenologic examinations were made of these cases, thereafter, and that it was not until the seventh month that both showed definitive improvement, radiologically. A third case followed a similar clinical course, but showed improvement after the fifth month. The opinion of both Doctors Amberson and Jones was that the combination of bed rest and hyperalimentation did bring about a greater improvement than was anticipated, in view of the advanced stage of the disease.

It may also be stated, here, that in not one of the three cases did the total plasma proteins rise significantly with the rise in body weight. This may mean either injury to the plasma protein regenerative mechanism, or that the particular hydrolysate used is a poor plasma protein regenerant. The latter supposition seems to be more correct, because Chow showed that, with this particular hydrolysate, he was unable to produce satisfactory regeneration of protein on depleted dogs.[20] Whether the use of a hydrolysate with demonstrated ability to regenerate plasma proteins more rapidly would bring about an accelerated improvement in the radiologic picture, remains to be tested.

Clinical Malnutrition

In a previous report,[19] two cases of clinical malnutrition were presented, one having lost 22 kilograms of weight, following a poorly performed Polya gastrectomy, and the other having lost 20 kilograms in six months, as a result of an undiagnosed partial intestinal obstruction. The first one was given protein hydrolysates, representing .75 grams of nitrogen/KBW, and the other, .883 grams N/KBW. It took almost four weeks to restore both to working condition, at which time their body weights had come back to within a few kilograms of normal. It was stated, then, that this period of rehabilitation was several times faster than is usually the case, when natural food is used. It was also remarked that the phenomenon of a decrease in body weight, in the face of a markedly positive nitrogen balance, might have been due to the elimination of an excess of extracellular fluids. This remark was

prompted by the determination of the extracellular fluid volume (thiocyanate space), along with blood volume and other values, which had been made in the second case, but was not reported, because it was the only case in which these determinations had been made.

Since then, two more such cases have been studied in a similar way, determining the total blood volume, the thiocyanate space, the plasma volume, and calculating from these and the hematocrit and the plasma protein level, respectively, the total red cell mass and the total circulating plasma proteins. The results of the study appear in TABLE 2.

TABLE 2

BLOOD ELEMENT AND FLUID COMPARTMENT STUDIES IN THREE CASES OF CLINICAL MALNUTRITION

TBV = Total Blood Volume; PV = Total Plasma Volume; TRCM = Total Red Cell Mass (erythron) = TBV x hematocrit; TPP = Total Plasma Proteins = P V x Plasma Proteins gms%; TEx-F = Total Extracellular Fluid (thiocyanate space); PP = Plasma Proteins (grams%).

Name	Date 1945	Weight Kgm	Type feeding N. intake gm/KBW	TBV L	P V L	Hematocrit %	TR CM L	PP gm%	TPP gm	TEx-F L	Ex-Fl cc/ Kgm.
G.D.*		74.09		6.67	3.7	45	3.00	6.5	240	14.82	200
	6/28	54.08	"A"	4.38	3.05	31	1.33	5.36	162	23.81	443
	7/9	59.5	.833 gm N	5.38	3.23	40	2.15	6.73	218	25.25	424
	7/24	65.9	per KBW	5.82	3.43	42	2.39	7.01	241	24.73	376
S.T.*		70.9		6.38	3.55	45	2.83	6 5	231	14.18	200
	5/3	56.4	"B"	3.84	2.44	43	1.4	5.42	174	18.21	323
	5/9	54.1	.6 gm N	3.93	2 37	40	1.56	5.56	178	16.67	308
	5/24	57.0	per KBW	4.21	2.59	38	1.62	6.09	168	19.19	335
	8/6	66.1		4.57	2 25	45	2.32	5.74	137	13.80	209
F.K.*		59.0		5 31	2.92	45	2 39	6 5	192	11.6	200
	9/25	40	"C"	3.10	2.01	35	1 09	5 92	119	13.89	347
	10 3	43.5	.575 gm N	2.97	2 05	31	0 92	5.58	114	10.0	230
	10/28	50.5	per KBW	3.59	2.16	40	1 43	5.78	124	11.5	228

* The figures in the same horizontal lines as the initials are values estimated from the body weight, which is also estimated.

The figures in the same horizontal column as the initials are estimated figures of body weight, of total blood volume (TBV), using the factor of 90 cc. per kilogram body weight; of plasma volume (PV), using the factor of 50 cc. per KBW; the total red cell mass (TRCM), by multiplying the total blood volume by the hematocrit; and the total circulating plasma protein (TPP), by multiplying the plasma volume

by the plasma protein level; the thiocyanate space, by multiplying the body weight in kilos by 200. This last item appears in the last column as a normal figure, with which to compare the determined values. The first four vertical columns represent the initials of the patient, the date, the weight in kilograms, and the type of protein hydrolysate fed, including the dose of nitrogen per kilogram body weight. The titles of the other columns are explained in the upper portion of the table.

It will be seen from the table that the initial weight of G.D. was 54.08 K; of S.T., 56.4 K; and of F.K., 40 K.; a loss, respectively, from the estimated pre-illness weight, of 27%, 20.4%, and 32%. The initial total blood volumes had dropped 34.3%, 40%, and 41% of their estimated pre-illness levels. The plasma volume, on the other hand, had dropped only 17%, 31%, and 31%, while the total red cell mass, by 56%, 50%, and 55%. The total plasma protein dropped 32.5%, 25%, and 38%. Thus, among these values, the total red cell mass seemed to have suffered most, the total plasma protein next, while the total blood volume and the plasma volume, least. The extracellular fluids, measured by the thiocyanate method, contrariwise, registered a significant rise; namely, 120%, 61%, and 73.5%, respectively.

At this point, it is interesting to compare values for the total red cell mass and total plasma proteins with the hematocrit and the plasma protein levels. It will appear from the table that, while the total red cell mass of G.D. had dropped from the estimated pre-illness level of 3 liters to 1.33 liters, a drop of 56%, the hematocrit dropped from 45% to 31%, or a drop of only 31%. The total plasma protein fell from 240 grams to 162 grams, a fall of 32.5%, while the plasma protein level fell only from 6.5 to 5.36 grams, or a fall of only 17.5%. Similar discrepancies were also found with S.T. and F.K. In the former, the total red cell mass fell 51%, and total plasma proteins, 24.7%; while the hematocrit and the plasma protein level fell only 4.4% and 16.6%, respectively. In F.K, similarly, the total red cell mass fell 59%, while the hematocrit fell only 22%; the total plasma proteins fell 38%, while the plasma protein level, only 8.9%.

Thus, the methods we have relied on, clinically, for information as to whether or not the number of red cells and the plasma proteins are normal (namely, the red blood count and the plasma protein level) are not as accurate as we are in the habit of thinking. This is probably true of the hemoglobin levels. A fairly well-developed oligemia and hypoproteinemia can be present, before they are detected by our relative methods of measurement. These hidden deficiencies could well

be the underlying causes of the depreciation of physical fitness, following operation and disease.

Three different types of protein hydrolysates were used in the rehabilitation of these three cases. The level of feeding was, unfortunately, not uniform. The body weight rose gradually during the hyperalimentation regimen, in G.D. and F.K., but in S.T., a case of pulmonary tuberculosis, it registered a fall first, despite a markedly positive nitrogen balance. The final weights rose 22%, in 26 days, in G.D.; 17%, in 93 days, in S.T.; and 26%, in 33 days, in F.K. Because of the fact that S.T. was a case of pulmonary tuberculosis, it is not fair to attribute this slow rise to the type of protein hydrolysate used.

During the regimen, the total blood volume rose 33%, 19%, and 14%, respectively, while the plasma volume rose 11% in G.D., fell 8% in S.T., and rose 7% in F.K. The total red cell mass rose almost 80% in G.D., and 31% in F.K., while it fell 21% in S.T. The values for the extracellular fluids, as indicated by the thiocyanate space, underwent a rise in G.D., and a fall in both S.T. and F.K.; but, in all three, there was a consistent fall in the volume of extracellular fluid per kilogram body weight. It thus appears that, when a patient loses nitrogen, he develops a tendency to store fluids in the extracellular space, but that, when he begins to store nitrogen, this excess fluid is eliminated, which, in some cases, accounts for a preliminary loss in body weight, in the face of a markedly positive nitrogen balance.

There does not seem to be a parallel between the weight loss and the fall of the total blood volume, plasma volume, total red cell mass, or total plasma proteins. Whether the lag of one value behind the other is due to the more rapid withdrawal of one or more amino acids than of others, is not clear. Nor is there a parallel in the rise of these values, when positive nitrogen balance is re-established. Whether this lag is due to injury to one or more of the specific mechanisms involved in the formation of these blood elements, is also a matter to be determined.

Effect of Hyperalimentation on Kidney Function

There has been in existence for a long time, now, the opinion that high protein diet is deleterious to the kidney. This has been disproved by the experience of Stefansson with his party of arctic explorers. Some recent experimental work by Addis,[21] showing that, when the kidney tissue of normal rats is reduced by surgical removal, until a third is left, the ingestion of a high protein diet precipitates uremia, seems to strengthen the former opinion.

A piece of work done in my laboratory, in which Dr. C. Y. Chu and Dr. C. S. Huang participated, is presented in TABLE 3. It is seen, from

TABLE 3

UREA CLEARANCE OF CASES ON PROLONGED HIGH NITROGEN FEEDING

Name	Clinical condition	N. Intake gm/KBW	Av. N. output	No. days avgd.	Time on high N. feedings prior to test	Urea clearance
G.D.	Malnutrition	.883	9 23	1st 10	1 month	74 8
S.T.	Tbc	.61	25.59	7th-24th	4 months, 4 days	79
K.P.	Tbc	.716	28.69	1st-24th	5 months, 11 days	84 8
C.Y.	Tbc	.8	32.94	7th-24th	5 months, 11 days	73

this table, that the ingestion of a high nitrogenous diet, over periods lasting from four weeks to five and a half months, failed to affect the urea clearance, nor did it cause albuminuria or any other detectably abnormal manifestations. How a nitrogenous intake will affect an individual with renal pathology, may be a different matter

Further Comments on the Clinical Use of Protein Hydrolysates

The main reasons for the use of protein hydrolysates have been set forth in the opening paragraphs of this work. A number of minor considerations may now be enumerated. Protein hydrolysates may be given in smaller bulk than a corresponding amount of natural food, and for this reason, more of the former can be given than of the latter. Being of uniform composition, it is easier to calculate the intake. It is easily dispensed. Given as a drug, it circumvents the vagaries of the patient's appetite and the administrative difficulty of prescribing a high protein diet in the form of natural food. The physician and nurse have first-hand quantitative knowledge of the patient's daily intake.

A great deal has been said of the unsavory taste of protein hydrolysates. This matter of taste is a relative one. Patients who need it and are told of the necessity of taking it, manage to down it. It needs some pressure, on the part of the attending physician, but certainly less than that required in obtaining the consent of the patient for a surgical operation. Moreover, the matter of taste is a chemical problem that is on the way to being solved.

The question of acceptability, however, does not depend entirely on

the taste alone. There are forms which provoke abdominal distress, flatulence, diarrhea, and delayed vomiting. This train of symptoms has some relationship to the bacterial count, particularly if putrefactive bacteria are involved. In a former paper,[19] I stated that the bacterial count should not be over 200 per gram of the material. In our present view, this statement is somewhat premature. The number of viable bacteria, at any one stage, is not an index of the bacterial history of a product, and hydrolysates having higher bacterial counts of 1000 have been given, without provoking symptoms, provided the bacteria are not of the putrefactive type. A test has yet to be devised to cover this problem of bacterial contamination.

All things remaining equal, a soluble product is more acceptable than an insoluble one. The salt content must be such as not to cause an over-dosage of any particular salt, when given in hyperalimentation levels. The product which can be tolerated in largest quantities is more useful.

There are now enough hydrolysates on the market to make therapeutic comparisons. Using the critical zone of nitrogen intake of herniotomy patients, we have found some hydrolysates to have the critical zone of .184 grams nitrogen per kilogram body weight, while others may be as high as .35 grams nitrogen per kilogram body weight. Manifestly, the former hydrolysates are nutritively superior to the latter.

Finally, if the lack of parallelism between body weight and total blood volume, total red cell mass and total plasma proteins, in malnourished subjects, is due not to injury of some specific mechanism, but to the amino acid compositions of the hydrolysates used, then we have found a method of testing what ideal mixture of amino acids would be required to restore these various body values in a parallel way, instead of having one lagging behind the others. This would make the use of the amino acids an epistomologically more productive avenue of study than that offered by natural food.

BIBLIOGRAPHY

1. **Cuthbertson, D. P.**
1942. Post-shock metabolic response. Lancet **1**: 433.
2. **Best, Charles H., & Norman Burke Taylor**
1939. The Physiological Basis of Medical Practice. Williams & Wilkins Co. Baltimore. (Sec. ed.)
3. **Albright, F.**
1942–43. Cushing's Syndrome, Harvey Lectures **1942-43**: 168
4. **Browne, J. S. L.**
Quoted by Albright (cf. 3.).

5. **Lund, C. C., & S. M. Levenson**
1945. Proteins in surgery. J. A. M. A. **128**: 95.
6. **Howard, J. E., J. Winternitz, W. Parson, R. S. Brigham, Jr., & H. Eisenberg**
1944. Studies on fracture convalescence: influence of diet on post-traumatic nitrogen deficit exhibited by fracture patients. Bull. Johns Hopkins Hosp. **75**: 209.
7. **Howard, J. E., E. Parson, K. E. Stern, H. Eisenberg, & V. Reidt**
1944. Studies on fracture convalescence: nitrogen metabolism after fracture and skeletal operations in healthy males. Bull. Johns Hopkins Hosp. **75**: 156.
8. **Howard, J. E.**
1945. Protein metabolism during convalescence after trauma. Arch. Surg. **50**: 166.
9. **Grossman, C. M., T. S. Sappington, B. A. Burrows, P. H. Lavietes, & J. P. Peters**
1945. Nitrogen metabolism in acute infections. J. Clin. Invest. **24**: 523.
10. **Co Tui, A. M. Wright, J. H. Mulholland, E. S. Breed, I. Barcham, & D. Gould**
1945. Studies in surgical convalescence II: A preliminary study of the nitrogen loss in exudates in surgical conditions. Ann. Surg. **121**: 223.
11. **Chittenden, R. H.**
1904. Physiological Economy in Nutrition. Frederick A. Stokes Co. New York.
12. **Chittenden, R. H.**
1907. The Nutrition of Man. Frederick A. Stokes Co. New York.
13. **Sherman, H. S., L. H. Gillett, & E. Osterley**
1920. J. Biol. Chem. **41**: 97.
14. **National Research Council**
1943. Recommended dietary allowances. Reprint and circular series **115**.
15. **Co Tui**
1945. The value of protein and its chemical components (amino acids) in surgical repair. Bull. N. Y. Acad. Med. **21**: 631.
16. **Elman, R.**
1940. Parenteral replacement of protein with amino acid of hydrolyzed casein. Ann. Surg. **112**: 594.
17. **Cox, W. M., Jr., & A. J. Mueller**
1939. Proc. Soc. Exp. Biol. & Med. **42**: 658.
18. **Co Tui, A. M. Wright, J. H. Mulholland, T. Galvin, I. Barcham, & G. R. Gerst**
1945. The hyperalimentation treatment of peptic ulcer with amino acids (protein hydrolysate) and dextri-maltose. Gastroenterol. **5**: 5.
19. **Co Tui *et al.***
1946. Some clinical aspects of protein nutrition. J. Am. Diet. Assoc. **22**: 97.
20. **Chow, Bacon F.**
Personal communication.
21. **Addis, T.**
Personal communication.

GENERAL DISCUSSION

Dr. John H. Mulholland (*Bellevue Hospital, and New York University College of Medicine, New York, N. Y.*):

From the clinical surgeon's viewpoint, no field of investigation by chemists and physiologists is more timely than protein nutrition. Many of the basic findings have direct application in the sick patient. As the field of surgery broadens itself to the correction of non-urgent disturbances, the operations become purely elective

and frequently extensive. The margin of safety in surgical risk must be extended, in parallel. This margin of safety is best determined by a measurement of the protein, or amino acid factors, in that general nutritional state.

Recent experiences in war surgery have been impressive. The wonderful response of a well-fed, well-conditioned American soldier to extensive wounds and operations has been attributed to penicillin, the sulfonamides, the use of whole blood, improved and earlier surgery, prompt transportation, and other factors. Frequently, the most important element of surgical success was forgotten. These soldiers had been trained, fed, and nourished for one to two years. Their superb condition gave surgeons a margin on which to work. In the later stages of the European campaign, it was a common experience to observe that, after extended field duties on K rations, the secondary wound closures and skin grafts were not so successful, and penicillin, blood, surgery, and transportation were used as before. Then, too, experience with liberated prisoners, liberated persons, and prisoners of war was eloquent testimony that a long period of malnutrition rendered all other advances less effective. Protein hydrolysates were not available, but adequate diet, plasma, and blood were not sufficient to bring about a *speedy* return to a well nourished state.

It is noteworthy, for the surgeon, that the clinical use of protein hydrolysates has been strikingly effective in the complications of diseases manifested by ulceration or poor healing powers in surface wounds. The integrity of the surfaces of the body seems to be closely linked with some phase of protein or amino acid metabolism. The essentiality of certain protein fractions is not unlike the vitamin requirement. The effectiveness of methionine, in the prevention and treatment of fatty liver, which occurred in soldiers who had been on K rations for long periods, was observed in the African theatre. This is certainly similar to vitamin deficiency replacement.

The trend of recent studies is not only to maintain positive nitrogen balance and plasma levels, but also to determine the specific properties of amino acids and the qualitative composition of proteins. The complexity of these problems, in this vast, labile, and dynamic chemical maze, is almost beyond the imaginative powers of a clinician. His service is to observe and evaluate controlled measures in the human.

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

VOLUME XLVII, ART. 4. PAGES 375–602

DECEMBER 15, 1946

THE PHYSICO-CHEMICAL MECHANISM OF NERVE ACTIVITY*

BY

DAVID NACHMANSOHN, CHARLES M. BERRY, OSCAR BODANSKY, FRANK BRINK, JR., DETLEV W. BRONK, M. VERTNER BROWN, C. W. COATES, R. T. COX, J. C. ECCLES, ALFRED FESSARD, J. F. FULTON, R. W. GERARD, ALFRED GILMAN, D. E. GREEN, JOSEPH C. HINSEY, RUDOLF HÖBER, MARTIN G. LARRABEE, AND TRACY J. PUTNAM

CONTENTS

	PAGE
INTRODUCTION TO THE CONFERENCE ON NERVE ACTIVITY. BY TRACY J. PUTNAM	377
THE MEMBRANE THEORY. BY RUDOLF HÖBER	381
CHEMICAL MECHANISM OF NERVE ACTIVITY. BY DAVID NACHMANSOHN	395
AN ELECTRICAL HYPOTHESIS OF SYNAPTIC AND NEURO-MUSCULAR TRANSMISSION. BY J. C. ECCLES	429
CHEMICAL EXCITATION OF NERVE. BY FRANK BRINK, JR., DETLEV W. BRONK, AND MARTIN G. LARRABEE	457
ELECTRICAL CHARACTERISTICS OF ELECTRIC TISSUE. BY R. T. COX, C. W. COATES, AND M. VERTNER BROWN	487
SOME BASIC ASPECTS OF THE ACTIVITY OF ELECTRIC PLATES. BY A. FESSARD	501
PHYSIOLOGICAL FUNCTION FROM THE STANDPOINT OF ENZYME CHEMISTRY. BY D. E. GREEN	515
CHOLINESTERASE. BY OSCAR BODANSKY	521
THE EFFECTS OF DRUGS ON NERVE ACTIVITY. BY ALFRED GILMAN	549
THE RECOVERY OF DIAMETER AND IMPULSE CONDUCTION IN REGENERATING NERVE FIBERS. BY CHARLES M. BERRY AND JOSEPH C. HINSEY	559
NERVE METABOLISM AND FUNCTION. BY R. W. GERARD	575
CONCLUSION. BY J. F. FULTON	601

* This series of papers is the result of a Conference on The Physico-Chemical Mechanism of Nerve Activity held by the Section of Biology of The New York Academy of Sciences, February 8 and 9, 1946.

Publication made possible through a grant from the Conference Publications Revolving Fund.

COPYRIGHT 1946

BY

THE NEW YORK ACADEMY OF SCIENCES

INTRODUCTION TO THE CONFERENCE ON NERVE ACTIVITY

By Tracy J. Putnam

Columbia University, New York, N. Y.

It is my pleasant duty to open a conference which is, in several respects, historic.

In the first place, this is, as far as I am aware, the first truly international conference on a purely physiologic or, in the broad sense, medical subject, since the beginning of the war. I see, in the audience, five continents represented, and I can reproach the program committee and myself only for failing to arrange for a delegate from Africa. It is my special privilege to welcome Dr. Eccles from Australia, Dr. Feng from China, Drs. Fessard, Couteaux, and Bugnard from France, and Dr. Bremer from Belgium, who have come so far for the special purpose of taking part in this symposium.

The fact that we are all met here from various corners of the earth to discuss problems of pure science with a humanitarian purpose symbolizes, it seems to me, the hope of this troubled world, the hope that civilization is beginning to recover from a desperate and destructive illness, which barely missed being fatal. A relapse might well be final. But here we are, ready, willing, and able to talk over some questions which are of great importance, but definitely non-political. It is to be expected that differences of opinion will arise, and facts will be hotly debated. I feel certain, nevertheless, that the members of this conference will be able to set an example for future international discussions, in agreeing on the criteria of truth and the means of arriving at an agreement on facts, with a broad tolerance towards possible means of interpretation. Let us hope that the United Nations Organization will take notice and be willing to learn.

This meeting opens a wholly new vista, in another sense, also. At the time when the German Army crossed the Polish border, in 1939, there seemed to be little hope of bridging the gap between the point of view that transmission of the nervous impulse was a purely electrical phenomenon, and on the other hand, the conception that the production of a specific chemical substance was the essential fact. The atmosphere of the war was not in the least conducive to placid scientific work;

Linlithgow Library.

but here, six years have passed, and behold! the whole subject has become suddenly clearer. Our new insight into the problem has not come easily, and I must pay special tribute to our colleagues in Europe and in China, who have had the courage to carry on their investigations in the face of enormous discouragements and practical difficulties, and sometimes even in secrecy and at the peril of their lives.

It is amazing that so much has been accomplished under such adverse circumstances. The details of the physiology of the individual elements of the nervous system have, in the past, seemed most obscure, and we have had to infer the outlines of the metabolic processes which occurred there from data gathered from other tissues. But it now seems safe to say that our picture of the metabolism and the mechanism of action of neurones is more complete than our knowledge of any other tissue, and the methods of study which were originally confined to neurophysiology are being extended to other physiologic problems.

It is particularly gratifying to me that a clinical neurologist should be permitted to open this meeting. Clinical neurology used to be considered a purely diagnostic specialty, a hopeless field of medicine, which consisted in little more than a meditation on disease. We are beginning now to be able to do a little more about the disorders of the nervous system, but we can make progress only as we possess insight. The physiologic methods of study which have been devised and applied by the distinguished scientists I see before me, and the facts they have elicited are, I am sure, the surest guide we possess to advances in therapeutic methods. This is a new chapter, not only in neurophysiology, but in pharmacology, clinical medicine, and, perhaps, even for the dark territories of psychiatry.

We are grateful to The New York Academy of Sciences, and especially to the executive secretary, Mrs. Miner, for having organized so efficiently and made possible this symposium. We were fortunate in having the support of the Rockefeller Foundation, and we express our gratitude to Dr. Lambert for his advice and active cooperation.

I should also like to thank very warmly Dr. Raymond Zwemer who, through his association with the State Department, helped us to overcome many difficulties.

A few words about the program. The purpose of the symposium is not to present recent data alone, but to give an opportunity for discussing some of the fundamental aspects and problems. There is so much to say that we have filled the program perhaps unduly full, and still have been unable to find time for many investigators whom we

should all enjoy hearing. We hope, however, there will be enough time left for discussion, if we abide closely by our schedule. The chemical aspect of the subject has been as much stressed as the physical. It seems that we shall all have to get accustomed to terms like enzymes and coenzymes, as well as to positive and negative phases, and Wedensky inhibition. At the end of the symposium, Dr. Gerard will try to integrate the different aspects which will be presented and discussed.

THE MEMBRANE THEORY

By Rudolf Höber

University of Pennsylvania, Philadelphia, Pennsylvania

The classical objects of the study of bioelectric phenomena are muscle and nerve. Resting, injury, and action potentials and currents are studied with both of them. On the basis of Wilhelm Ostwald's investigations upon the electric properties of artificial inorganic precipitation membranes (1890), the physiological membrane potentials have been looked upon as being special forms of Nernst concentration potentials; in other words, potentials arising when solutions of different electrolytes are separated by a membrane characterized by a more or less selective ion permeability. As it was from the beginning of the electrophysiological era, both nerve and muscle have been used for solving the basic problems, and information gained from one type is valuable for both. Therefore, although our object is primarily a discussion of physicochemical mechanisms of nerve activity, muscle potentials will also be treated.

The *Membrane Theory* was established, in 1902, by Bernstein, when he ascribed the EMF of the locally injured muscle fiber to selective permeability to potassium ions present inside the fiber in a considerably greater concentration than outside. He conceived of the action potential wave as a self-propagating depolarization by breakdown of this selective permeability. It was early assumed that this alteration is accompanied by chemical reactions.

THE ION PERMEABILITY OF THE RESTING MUSCLE AND NERVE

In immediate connection with Bernstein's theory, in order to become acquainted with the general properties of the selective ion permeable membranes, it seemed to me that one of the main tasks was to study the effect of local application of the neutral inorganic salts upon the resting potential of muscle (Höber[1]).

A. The Inorganic Cations

At first, the *alkali cations* only were varied, and potassium was found to produce the strongest negative pole, similar to the effect of

cutting the muscle, but often as a reversible process and obviously indicative of the greatest permeating power. The other alkali cations appeared to be less effective, in this order: potassium, rubidium, sodium, lithium; potassium and rubidium producing negativity, as compared to sodium, while lithium produces positivity. This was interpreted as being due to swelling or shrinking of hydrophilic colloids, which were assumed to be the chief constituents of the plasma membrane. Later, the same series was met by Michaelis with non-colloidal, rigid, dried collodion membranes. His findings seemed to fit in best with the concept of an ionic sieve. For, taking into consideration the shells of water dipoles around the ions, the effective ionic volume appears to be smallest with potassium, largest with lithium. But, as will be seen later, the differences are better correlated with adsorption, which, according to Gouy, Frumkin, and others, increases with decreasing hydration of the ions, potassium being most, lithium least, adsorbed.

The membrane theory postulates that, as in the case of the sufficiently dried collodion membrane, which is permeable only to potassium ion, strength and direction of the injury potential are dependent upon the ratio of potassium inside to potassium outside. In other words, the surface of muscle or nerve behaves as a potassium-electrode, potassium inside being constant and about 20 to 40 times greater than potassium outside. Therefore, by raising potassium outside, the EMF of an injured (cut) muscle should be decreased to zero, if potassium outside is equal to potassium inside, and its direction should be reversed, if potassium outside is greater than potassium inside. My own early experiments (1905) failed to show the reversal, because, in contrast to the rigid, dried collodion membrane, the ion selectivity of the plasma membrane is lost, due to its colloidal behavior: in other words, due to the swelling and disintegrating, even to the cytolyzing effect of higher potassium, especially after some lapse of time. However, the postulate of a reversal complies with recent observations of Hodgkin and Huxley,[2] and of Curtis and Cole,[3] in a particularly striking way. These authors, leaning upon Osterhout's[4] studies on the "impaled" giant plant cells (*Valonia*), pushed a microelectrode into the axoplasma of the giant nerve fiber of the squid, along its axis, so that its tip was placed just opposite to the outside electrode, and they thus measured the membrane potential directly across the wall. The potential was found, in the case of the squid nerve, to be, on an average, 50 mV. Then, upon raising potassium outside to about 18 times normal, the resting potential was decreased to zero, and upon raising it about 40 times normal, a reversal of 15 mV was observed. The corresponding

procedure has been applied to single frog muscle fibers by Gerard, Carlson, and Graham.[5]

B. The Inorganic Anions

The colloidal behavior of the physiological membranes, further, is brought into evidence by the effect of *the inorganic anions.* In general, with regard to their role in physiology, the anions are less powerful than the cations. This is due to the prevalent negative charge of the colloidal aggregates, which repel the anions. Locally applied to nerve or muscle, the resting potential reveals the following anion series: thiocyanide, iodide, bromide and chloride, sulfate, with thiocyanide ion producing a *positive* pole. In other words, a reversed injury potential occurs as a consequence of the anion adsorption on the membrane, which is greatest with thiocyanide, smallest with sulfate. This reversal is an important point to be kept in mind for later discussion. However, after some lapse of time or after applying the salts in stronger solution, the anion series is reversed, thiocyanide and iodide forming a *negative* pole, thus resembling the effect of potassium ion. This, again, is significant, as due to a loosening, softening, and subsequent disintegrating action upon the hydrophilic colloidal membrane.[1, 6]

C. Organic Anions

These effects are related to those of a large group of *organic anions:* for instance, those of higher fatty acids starting with the 8-carbon atom chain, *i.e.*, caprylic acid. These ions have a nonpolar-polar structure, the nonpolar, or organophilic and hydrophobic, part of the anion, mainly due to the adsorption affinities of the alkyl radicals, attaching to the organic material, *e.g.*, to the particles of a Langmuir surface film of protein; the other polar or hydrophilic part, due to the carboxyl radicals with their cloud of water dipoles, anchoring in the water. It appears that a pull towards the water can be exhibited upon the organic material. The pull is stronger or weaker, corresponding to the relative hydroaffinity, so that colloidal particles composed of a variety of molecules, like hemoglobin (which means, hem plus globin), visual purple, or chloroplastin, can be torn to pieces, a process termed denaturation, solubilization, or detergency. On the other hand, coiled peptide chains can be uncoiled, and organic architectures of great complexity, such as a plasma membrane, can be loosened by the adsorptive pull, which has the effect of abolishing reversibly the selective ion permeability of the membrane and starting irreversible cytolysis.[7] For example, sodium caprylate is applied locally to a muscle. The re-

sult of a very weak solution is simply a reversed resting potential, due to the anion attachment to the pores of the membrane; the effect of a stronger solution is a regular resting potential, due to reversible loosening; while the result of a still stronger solution is an irreversible disintegration, *i.e.*, an injury potential.

D. Inorganic Plurivalent Cations

The occurrence of the well-known *cation antagonism* is another indication of the prevalence of anionic colloids in cell structure. An imbalance between the monovalent and the plurivalent cations shows up in numerous observations upon animal and plant cells, among others by alteration of their electrical properties, as, for example, ohmic resistance, or conductance, or excitability. Preponderance of monovalent cations (sodium, potassium) is bound up in muscle and nerve with loss of normal selective cation permeability, due to increased hydration. However, this is compensated for by the consolidating effect of plurivalent cations, like calcium, strontium, barium, cobalt, manganese, nickel (Höber[8]).

THE CHANGE OF ION PERMEABILITY BY DC

Let us turn now to the old and complex phenomenon of *electrotonus*. It comprises a multitude of alterations of cell responses, effected by direct current and having their origin in changes of membrane polarization. Especially well-known are the changes of excitability of nerve and muscle. Excitability is diminished at the anode and increased at the cathode, except that, beyond a certain current strength, the increase turns to a decrease, the so-called cathodic depression. Anelectrotonus and catelectrotonus are tied up with changes of resistance. By placing one electrode on an intact spot of the excised nerve, the other on a crushed end, the resistance is raised at the anode, diminished at the cathode. Consequently, while sending alternating current through the preparation instead of direct current, a *rectifier effect* appears. These and other observations can be explained, partly on the basis of ion distribution between the inside and the outside of the membranes, partly by taking again into account the colloidal properties of the wall of the natural membranes. As to the first point, according to Bear and Schmitt,[9] Cowan,[10] Fenn and others,[11] and Webb and Young,[2] the axoplasm of the giant nerve fiber of the squid, for example, contains 4 to 5 times more inorganic cations (mainly potassium) than anions, and 18 times more potassium than the blood. Be-

sides chloride, there are in the axoplasm small concentrations of phosphate, sulfate, and lactate, but rather large amounts of organic anions of low mobility, possibly the anions of amino acids. These conditions are roughly reproduced in model experiments of Labes[12] and Ebbecke.[13] A membrane core-conductor is formed by a collodion tube, with pores wide enough to allow cations and anions to pass the wall. The tube is filled with a solution of potassium phosphate and is packed in gauze which has been wetted with a solution of sodium chloride. One electrode is placed inside, another outside. If direct current passes the membrane, a smaller resistance is encountered, when the current goes from within outwards, than when it goes in the opposite direction. The reason is that, with the outgoing current, the faster potassium inside and chloride outside are swept into the membrane and travel, there, with greater velocity than sodium outside and phosphate inside, being driven by an ingoing current. If, instead of collodium, hydrophilic and negatively charged colloids are the membrane constituents, as they actually are under most physiological conditions, then additional swelling and increasing dispersion occur at the cathode, as well as shrinking at the anode. Consequently, the polarizability of the membrane falls at the cathode and rises at the anode. Swelling causes in natural objects, such as muscle and nerve, greater excitability at the cathode, but as the current strength rises more and more, the higher excitability turns over to inexcitability, or, in other words, to cathodic depression. More specifically, according to Blinks,[14] the membrane polarization of a giant cell of the fresh water alga *Nitella,* comparable to nerve or muscle with its thread-like shape, drops down to zero, if exposed to the swelling effect of potassium chloride in sufficiently high concentration, and the cell does not respond any more to otherwise effective stimuli. However, by applying. locally, an anode of rising strength to the depolarized *Nitella* cell, above a certain threshold value, the polarizability is restored, and a normal action potential can be elicited upon stimulation. Alternatively, with respect to cation antagonism, after excitability of a nerve has been suppressed by calcium, this stiffening effect is cancelled by the softening influence of a cathode, as shown by Woronzow,[15] and more recently by Guttman and Cole.[16]

We turn, now, to the discussion of natural *changes of ion permeability during action.* It has been accepted, for more than 40 years, that depolarization, which is brought about by injury, compares essentially to depolarization accompanying excitation, as indicated by the "negative variation" of du Bois-Reymond. The negativity wave, therefore, is interpreted as a "breakdown of the membrane," by which the selec-

tive cation permeability is abolished. Correspondingly, it has been assumed that, during excitation, the potential fall is as great as is the resting potential, measured at best with the impaled nerve or muscle. But this is not true. Impedance measurements have shown the resistance to persist, to some extent, during excitation (Curtis and Cole). In other words, the resting potential could be expected to be larger than the action potential. However, the contrary is true. Hodgkin and Huxley,[2] and Curtis and Cole,[1] inserting a microelectrode into the axon, detected the potential change, during activity, to be even larger than that due to injury. For example, in the experiments of Curtis and Cole, the resting potential average is 51 mV, the action potential 108 mV.

Before discussing this interesting situation, attention will be turned briefly to a special problem. The word, breakdown, suggests leakage, and for this reason, activity could be expected to be accompanied by *leakage*, especially from the large surplus of well-diffusible potassium normally retained in the axoplasm. However, such an escape from frog nerve, though often investigated, is doubtful, except following very prolonged stimulation (for example, 60 stimuli per second, for 1–3 hours, in the experiments of Arnett and Wilde, with Fenn).[17] However, this may be accounted for, by assuming that only a very small area of the surface of a myelinated nerve, the Ranvier nodes, is available for diffusion. This can be correlated with the experiments of Cole and Curtis,[18] regarding impedance and membrane capacity of the squid nerve. Notwithstanding the fact that, during excitation, the resistance of the squid nerve falls off from 1000 ohm/square-cm. to only 25 ohm/square-cm., not more than 2% of the area is involved in the increase of permeability. This means that the remainder, about 98%, would be inactive. Another point is the fact that the state of excitation, in general, lasts only a very short time, measured in milliseconds. Very slowly reacting cells, therefore, may offer a greater chance to detect an ion escape. As a matter of fact, the conductivity of the water on the outside of the surface of a *Nitella* cell rises perceptibly, after several excitation waves have passed the slowly responding object, the excitation time being measured in tenths of a second (Cole and Curtis). Since depolarization is followed by repolarization, the question arises, whether and how the ions which escape through the leaky membrane are recovered. It becomes increasingly clear that, in one way or the other, energy is utilized for this purpose. In other words, the physiological membranes are more than labile structures. Rather, they are, or can be, acting machineries.

For example: According to Furusawa, Feng, and Shanes and Brown,[19] during anoxia the polarization of crab nerve and its excitability fall off reversibly, but seem to be maintained in the presence of phosphopyruvic acid, adenosintriphosphate, and thiamin; in other words, by establishing the normal glycolytic cycle. According to Hoagland and Davis,[20] *Nitella* cells in the dark lose their intracellular chloride ions, through the protoplasmic wall, into the surrounding water and recapture them during exposure to light. Furthermore, according to J. E. Harris,[21] potassium ion gets lost from human erythrocytes at low temperature, but re-enters at room temperature, after addition of glucose.

REVERSAL OF THE NORMAL ACTION POTENTIAL

I now come back to the lately-discovered fact, already mentioned, that the potential change during action does not equal the resting potential in magnitude, as it was assumed for many years. Rather, by overshooting the zero line, as shown in FIGURE 1, the potential is

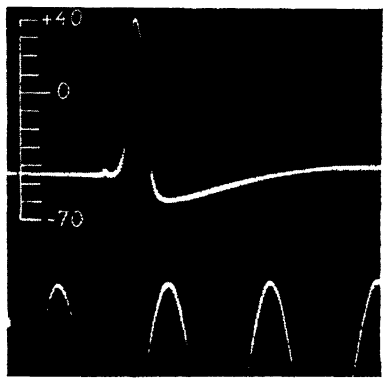


FIGURE 1. Potential of the internal electrode. The figure shows that the resting potential is —44 mV. During activity, the potential overshoots to the positive side, +40 mV, so that the action potential wave amounts to 84 mV. (**Hodgkin & Huxley**[4].)

momentarily reversed in sign, *the outside of the membrane becoming negative to the inside.* This reversal during passage of the impulse does not fit into the classical picture of the behavior of the active nerve membrane, and possibly indicates a special mechanism, which is superimposed to the mechanism of the customary excitation depolarization. FIGURE 2 depicts three conditions of the nerve membrane: (a) represents the normal polarization of a resting nerve membrane; (b) is indicative of the depolarized membrane, which, according to the ordinary view-

point, during activity is fairly equally permeable to cation and anion, to this extent resembling the situation in injury; while, in (c), an additional influence of organic anions has been taken into account, as has been discussed already by Hodgkin and Huxley.[2] These authors

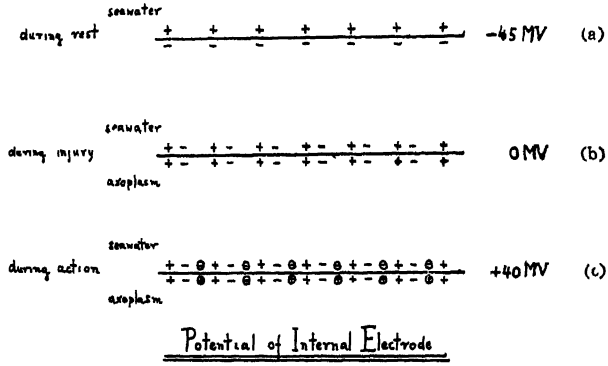


FIGURE 2 Diagrams illustrating the reversal of the nerve potential

have considered, among others, particularly the lactic acid anion, which, during activity, would penetrate the membrane from inside and produce a negativity outside. However, this hypothesis is rejected by Hodgkin and Huxley themselves, as it would be hard to imagine the concentration and the mobility in the membrane of the lactate ion as being sufficient. Instead, I would prefer to pay attention, especially, to the organic nonpolar-polar, hydrophobic-hydrophilic anions, already mentioned, which possibly can be assumed to be present in the nerve membrane, or, rather, to be liberated as the excitation wave travels along the fiber. As stated earlier, such anions, locally applied to the outside of a muscle, call forth a *reversed resting potential,* whereas, if they originated during excitation inside, they would call forth a *reversed action potential,* due to the fact that the adsorption forces would draw these anions into the porous membrane, as shown in FIGURE 2(c). Such a reversed resting potential has been found with the salts of higher fatty acids, alkyl sulfates, aryl sulfonates, and others. These experiments should be extended to nerves, especially to single nerve fibers like that of the squid, for the following reason:

Nonpolar-polar anions are abundantly preformed in the molecules of lipoids of the nervous system, chiefly in phospholipids and cerebrosides. Among their split products, the nonpolar-polar character is especially pronounced in the anions of fatty acids with long carbon atom chains, and, according to Langmuir and Adam, particularly in fatty acids

with one to three double bonds, *e.g.*, in oleic, linolic, linoleic, arachidonic, nervonic, and oxynervonic acids. These long carbon atom chains of the lipoids, lecithin, kephalin, sphingomyelin, and cerebrosides seem to be existent, not only in the massive sheath of the myelinated nerve fibers, but, according to Young and Francis Schmitt,[22] also in the unmyelinated fibers of crabs and cephalopods (for instance, the squid nerve), where the thickness of the sheath has been found to be as small as one per cent of the diameter of the axon, *i.e.*, about 5μ, compared to 25 per cent of the diameter in vertebrates, as shown by Pumphrey and Young (PLATE 1). In the sheath of the unmyelinated fibers, the lipoids, though often not demonstrable by the customary staining with osmium tetroxide, can be detected with polarization optics (Bear and F. O. Schmitt[23]).

Now, looking upon the excitatory process from the standpoint of the old "Strömchen theory" of Hermann,[24] it is at the boundary between the stimulated altered and the adjacent, unaltered region that small local circuit currents arise, flowing *out* of the unaltered region, which then secondarily gets altered as in a catelectrotonus, and flowing *in* at the originally stimulated region, which, thus, is inactivated as in an anelectrotonus. Catelectrotonus, however, as mentioned before, means softening the colloidal membrane and dispersing its structural aggregates by way of potassium and chloride ions and depolarization of the normal resting membrane. Anelectrotonus, on the other hand, means condensation and re-polarization. The aforementioned increase of concentration of potassium, which happens to be produced in the membrane by the outflowing current, may then serve to liberate in the nerve membrane, directly, some of the nonpolar-polar anions, as, according to the well-known studies of G. L. Brown and W. Feldberg,[25] acetylcholine is liberated by even a very small surplus of potassium (amounting to not more than 0.01 per cent) in the perfused ganglion cells, where it normally is fixed in a nondiffusible state However, the mechanism of this release is by no means clearer than that just suggested for the nonpolar-polar anions. Alternatively, the nonpolar-polar anions could possibly be liberated, indirectly, by an activation of lecithase A, an enzyme occurring in nerve tissue, which is known to set free the unsaturated, but not the saturated, fatty acids of the lipoid molecules.[26]

These are speculations, it is true. If, however, we refer them to the giant axon of the cephalopods, which was studied, in recent years, with most diversified and modern methods, it means that probably the alterations are bound up with the thin surface membrane which wraps up the voluminous column of axoplasm, and that this fine mem-

brane would be the site of a complex chemistry. Although the discussion of the chemical side of nerve activity is beyond the scope of this paper, I should like to conclude by turning to some interesting observations of von Muralt (1942), involving the appearance, during excitation, of a substance which may bring about the reversal of the membrane polarization.

When an excised frog sciatic is stimulated, electrically, at a certain frequency and simultaneously is dipped with a certain velocity into liquid air, several excitation waves must be caught and frozen along the nerve. When an extract of stimulated and unstimulated nerves, pulverized in the frozen state, is made up with eserinized frog Ringer or serum, it appears that, during excitation, the nerve has liberated minute amounts of *several* substances. One of these, by various tests, is identified as acetylcholine; and a second substance is concentrated in the foam of the extract, which, from this sign of surface activity, possibly indicates the presence of a nonpolar-polar substance, whereas, in the foam from an acetylcholine-eserin-serum solution as a control, the acetylcholine fails to show an accumulation. Recently, by the same freezing method, von Muralt has intercepted a third substance, thiamin, which possibly also is surface-inactive.[25] Certainly, these results are far from giving conclusive support to the concept that nonpolar-polar substances, detectable by their surface activity, have been liberated during excitation. Even if they were, the liberation may be of minor significance, considering the fact that, according to Hopkins and Huxley, and to Curtis and Cole, the resting potentials of the giant nerve fibers vary little from one experiment to another, in contrast to a wide variability appearing in the size of their action potentials. In any case, this group of observations emphasizes the urgent need to extend the study of chemical products, which are *directly* connected with nerve activity, beyond the demonstration of acetylcholine.

BIBLIOGRAPHY

1. **Höber, R.**
1905. Pflüg. Arch. ges. Physiol. **106**: 599.
2. **Hodgkin, A. L., & A. F. Huxley**
1939. Nature **144**: 710.
1945. J. Physiol. **104**: 176.
Webb, D. A., & J. Z. Young
1940. J. Physiol. **98**: 299.
3. **Curtis, H. J., & K. S. Cole**
1942. J. Cell. Comp. Physiol. **19**: 135.
4. **Osterhout, W. V., *et al.***
1927. J. Gen. Physiol. **11**: 193.

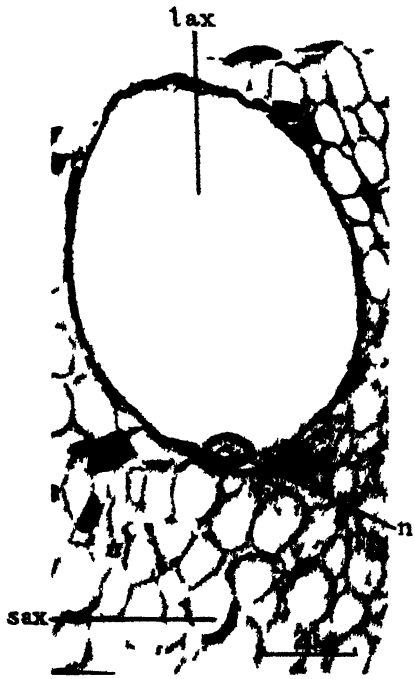
5. **Graham, J., G. R. Carlson, & R. W. Gerard**
1942. Fed. Proc. **1**: 31.
6. **Höber, R., M. Andersch, J. Höber, & B. Nebel**
1939. J. Cell. Comp. Physiol. **13**: 195.
7. **Höber, R.**
Unpublished Data.
8. **Höber, R.**
1917. Pflüg. Arch. ges. Physiol. **166**: 531.
9. **Bear, R. S., & F. O. Schmitt**
1939. J. Cell. Comp. Physiol. **14**: 205, 351.
10. **Cowan, S. L.**
1934. Proc. Roy. Soc. London B **115**: 216.
11. **Fenn, W. O., *et al.***
1934. Am. J. Physiol. **110**: 74.
12. **Labes, R., & H. Zain**
1926. Arch. Exp. Path. Pharm. **125.**
1927. Arch. Exp. Path. Pharm. **126.**
13. **Ebbecke, U.**
1933. Ergebn. Physiol. **35**: 756.
1922. Pflug. Arch. ges. Physiol. **195**: 555.
14. **Blinks, L. R.**
1936. J. Gen. Physiol. **20**: 229.
15. **Woronzow, D. S.**
1924–1929. Pflüg. Arch. ges. Physiol. **203, 207, 210, 216, 223.**
Höber, R., & H. Strohe
1929. Pflüg. Arch. ges. Physiol. **222**: 71.
16. **Guttman, R., & K. S. Cole**
1941. Proc. Soc. Exp. Biol. & Med. **48**: 293.
Guttman, R.
1940. J. Gen. Physiol. **23**: 343.
17. **Arnett, V., & S. Wilde**
1941. J. Neurophysiol. **4**: 572.
18. **Cole, K. S., & H. J. Curtis**
1939. J. Gen. Physiol. **22**: 649.
19. **Furusawa, K.**
1929. J. Physiol. **67**: 325.
Feng, T. P.
1932. J. Physiol. **76**: 477.
Shanes, A. M., & D. E. S. Brown
1942. J. Cell. Comp. Physiol. **19**: 1.
20. **Hoagland, D. R., & A. R. Davis**
1923. J. Gen. Physiol. **5**: 629.
21. **Harris, J. E.**
1941. J. Biol. Chem. **141**: 579.
22. **Young, J. Z.**
1936. Proc. Roy. Soc. London B **121**: 319.
Schmitt, F. O.
1936. Cold Spring Harbor Sympos. **4**: 7.
Pumphrey, R. J., & J. Z. Young
1938. J. Exp. Biol. **15**: 433.
23. **Bear, A. S., & F. O. Schmitt**
1937. J. Cell. Comp. Physiol. **9**: 275.
Young, J. Z.
1933. Cold Spring Harbor Sympos. **4**: 1.
24. **Lillie, R. S.**
1923. Protoplasmic Action and Nerve Action. Chicago Univ. Press. Chicago, Illinois.

25. **Brown, G. L., & W. Feldberg**
1936 J. Physiol. **86**: 290.
26. **Belfanti, S., A. Contardi, & A. Ercoli**
1936. Erg. Enzymforsch. **5**: 213.
Schmidt, G., B. Hershman, & S. J. Thannhauser
1945. J. Biol. Chem **161**: 523.
27. **von Muralt, A.**
1942. Pflug. Arch. ges. Physiol. **245**: 604.
28. **Liechti, A., A. von Muralt, & M. Reinert**
1943. Helvet. Physiol. Acta. **1**: 79.
von Muralt, A.
1945. Experientia **1**(5).

PLATE 1

PLATE 1

Large axon and small axons of stellar nerve of *Sepia officinalis* (Young[4])


lax
n
sax

CHEMICAL MECHANISM OF NERVE ACTIVITY

By David Nachmansohn*

Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, N. Y.

INTRODUCTION

The electrical signs of nervous action were, for a century, the only manifestations studied by neurophysiologists. But the function of a living cell cannot be conceived in purely physical terms. This was clearly expressed by Gasser, when he compared the electric spikes to the ticks of the clock, both being only signs of activity.[1] For a thorough understanding of the mechanism of nerve activity, a knowledge of the chemical reactions involved is essential. Biophysics and biochemistry are, consequently, of equal importance and inseparable in any attempt to solve the problem.

The special function of the nervous system is that of carrying messages from one distant point of the body to another. This process may be subdivided into three successive phases: First, a stimulus reaching a neuron has to initiate an impulse. Second, the impulse once initiated has to be propagated along the axon. Finally, the impulse arriving at the nerve ending has to be transmitted either to a second neuron or to an effector cell. Early in this century, T. R. Elliot had the idea that the third phase, namely, the transmission of the nerve impulse from the nerve ending to the effector cell, may be carried out by a chemical compound released from the nerve ending and acting directly on the second unit. Elliot suggested that adrenaline may be the transmitter of the impulse from the sympathetic nerve ending to the effector cell.[2] He based this idea on the similarity between the action of adrenaline and the effect of stimulation of sympathetic nerves on the effector organ. Similar ideas were advanced subsequently by Dixon and Howell.

In 1921, Otto Loewi found that, following vagus stimulation of an isolated frog's heart, a compound appeared in the perfusion fluid which, when transmitted to a second heart, produced an effect similar to that of vagus stimulation. Accepting the basic idea of Elliot, Otto Loewi concluded that this compound, which was later identified with acetyl-

* Most of the work described in this lecture has been supported by grants from the Josiah Macy, Jr., Foundation and the Dazian Foundation for Medical Research.

choline (ACh), is actually released from the nerve ending and acts on the heart cell directly.[3] The concept of "neurohumoral" transmission appeared enlightening in the case of the autonomic nerves and was widely accepted among physiologists.

In 1933, Dale tried to extend this idea of a "chemical mediator" of the nerve impulse to the neuromuscular junction and to the ganglionic synapse.[4] In this case, however, the theory encountered strong opposition. In addition to many contradictions and difficulties, summarized by Eccles,[5] there were two main objections. The first was the time factor. The transmission of nerve impulses across neuromuscular junctions and ganglionic synapses occurs in milliseconds. No evidence was available that the chemical process can occur at the high speed required, and Dale admitted this difficulty. The second objection was still more fundamental. According to leading neurophysiologists, the excitable properties of axon and cell body are basically the same. The electric signs of nervous action do not support the assumption that the transmission of the nerve impulse along the axon differs, fundamentally, from that across the synapse.

The idea of a chemical mediator, released at the nerve ending and acting directly on the second neuron, thus appeared to be unsatisfactory in many respects (Fulton[6]).

NEW APPROACH

Recognition of two features of nervous action is essential to an understanding of the problems and the difficulties involved: The high speed of the propagation of the impulse, and the smallness of the energy required. In medullated mammalian nerve, the impulse travels at the rate of 100 meters per second, and the energy required per impulse per gram is less than 1/10 of a millionth of a small calorie. The recording of such an event offers many difficulties, even with the use of specialized physical methods. Only in the last twenty years have really adequate instruments become available for the analysis of physical aspects of nervous function.

It is obvious that the study of the chemical reactions connected with an event of this kind must offer even more serious difficulties. No adequate methods are available for directly determining chemical compounds appearing in such minute amounts and for such short periods of time. There is, however, another possible approach. Nearly all chemical reactions in the living cell are effectuated by enzymes. The study of enzymes *in vitro* has elucidated many chemical reactions,

known to occur in living cells, which could not be followed by direct chemical determination of the compounds metabolized. Especially for an event occurring with such a high speed as the propagation of the nerve impulse, analysis of the enzyme systems involved appeared to be the most promising approach.

Enzyme studies alone are, however, not sufficient for the elucidation of a biological mechanism, since there are so many simultaneous enzymatic reactions in the complex system of the living cell. It is necessary to correlate enzyme activities with events in the intact cell recorded by physical methods. The most conspicuous example of such an approach is the development of muscle physiology. Through the pioneer work of A. V. Hill and O. Meyerhof, many physical and chemical changes have been correlated, and our concept of the mechanism of muscular contraction has, according to an expression of A. V. Hill, gone through a real "revolution."

The question of the role of ACh in the mechanism of nerve activity has been approached by the study of the enzyme systems involved in the formation and hydrolysis of the ester. On the basis of their behavior *in vitro*, the activities of the enzymes could be correlated in different ways with events in the living cell recorded by physical methods. The facts established show that the original theories of the role of ACh, and, more generally, the idea of "chemical mediation," have to be modified. There is a strong body of evidence that the release and removal of ACh is an intracellular process, occurring at points along the neuronal surface and directly associated with the nerve action potential. The agent, however, which transmits the impulse along the axon, as well as across the synapse, is the action potential.[6-9] Some of the most important features of these investigations may be briefly outlined.

I. CHOLINESTERASE

A. Time Factor

ACh is inactivated by the enzyme cholinesterase, which hydrolyzes the ester into choline and acetic acid. The first essential result of the studies of this enzyme has been the evidence of its high concentration in nerve tissue: Significant amounts of ACh may be split in milliseconds; that is, a period of time of the order required for the passage of a nerve impulse. Consequently, the potential rate of ACh metabolism is thus sufficiently high to permit the assumption that it parallels the rate of the electric changes and may, therefore, be directly connected with the nerve action potential.

The special case in which this problem of the time factor has been studied and received a satisfactory answer, is the frog's sartorius muscle.[10] A small fraction of this muscle is free of nerve endings. By determining the concentration of cholinesterase in this part of the muscle, in the part containing nerve endings, and in the nerve fibers, it is possible to calculate the concentration of cholinesterase at the motor end-plates. Since the number of end-plates in a frog's sartorius is known, the amount of ACh which may be split during one millisecond at a single motor end-plate can be calculated. This turns out to be 1.6×10^6 molecules of the ester. About one-third of the enzyme at the motor end-plate appears to be localized inside the nerve ending. On the assumption that one molecule of ACh covers about 20–50 square Å, the amount which may be hydrolyzed during one millisecond at one end-plate would cover a surface of 100–250 square microns.

A high concentration of cholinesterase, of an order of magnitude similar to that at motor end-plates, exists at all synapses, whether central or peripheral, mammalian or fish, vertebrate or invertebrate.[11] In mammalian brain, for instance, 10^{14} to 10^{15} molecules of ACh may be activated per gram of tissue during one millisecond. This corresponds to about 10–100 millions of square microns of neuronal surface.

These experiments removed one of the chief difficulties from the theory that ACh is involved in the transmission of nerve impulses. They established that the ester may be metabolized at the high speed required for a chemical reaction directly connected with such a rapid event.

The difference between synaptic region and fiber is, however, only quantitative. The concentration of cholinesterase is high everywhere in nerves, although it rises at the region of synapses.

B. Localization of Cholinesterase at the Neuronal Surface

The second essential feature is the localization of cholinesterase in the neuronal surface. Direct evidence for this localization has been offered with experiments on the giant axon of squid (*Loligo pealii*[12]). This axon has a diameter ranging from 0.5 to 1.0 mm. The axoplasm may be extruded and thus separated from the envelope. The envelope is formed of connective tissue, lipoid and plasma membrane. The axoplasm was found to be practically free of cholinesterase. The whole enzyme activity is in the envelope.

This exclusive localization of an enzyme in the neuronal surface has been found only in the case of cholinesterase. Respiratory enzymes are localized nearly completely in the axoplasm.[13] Bioelectric phe-

nomena occur at the surface. The high concentration of the enzyme at the surface suggested that ACh may be connected with conduction along the axon, as well as with transmission across the synapse. This view is consistent with the conclusion of neurophysiologists that the mechanism of these two events is fundamentally the same.

II. CORRELATION BETWEEN ENZYME ACTIVITY AND PHYSICAL EVENTS DURING NERVOUS FUNCTION

The high rate of ACh metabolism and the localization of the enzyme at the neuronal surface made possible the assumption that the ester is connected with the electrical manifestations of nerve activity. However, suggestive as these facts may be, observations on enzymes, as pointed out before, do not permit an interpretation of the actual role of the substrate. For an understanding of the precise function of an enzyme, its activity has to be connected with events in the living cell which, in the case of nerve, can only be recorded by physical means. Such a relationship has been established in three different ways.

A. Parallelism Between the Voltage of the Action Potential and Cholinesterase Activity

The first line of investigations in which a correlation between physical and chemical processes was obtained, was in experiments on the electric fish. It was found that the activity of cholinesterase in the electric organ parallels exactly the voltage of the action potential.

The powerful electric discharge in these organs is identical in nature with the nerve action potential of ordinary nerves (A. V. Hill[14]). The only distinction is the arrangement of the nervous elements, the electric plates in series. The potential difference developed by a single element is about 0.1 volt, which is the same order of magnitude as that found in ordinary nerves. In the species with the most powerful electric organ known, *Electrophorus electricus*, the so-called electric eel, several thousand elements are arranged in series from the head to the caudal end of the organ. Thus, the voltage of a discharge amounts to 400–600 volts, on the average, and, in some specimens, more than 800 volts have been observed. In *Torpedo*, another species with a powerful electric organ, the elements are arranged in a dorso-ventral direction. Since it is a flat fish, the number of plates in series usually does not exceed 400 to 500, and, consequently, the discharge is only 30 to 60 volts, on the average.

In 1937, an extraordinarily high concentration of cholinesterase was found in the strong electric organ of *Torpedo*. In the following year,

a similar high concentration was found in the electric organ of *Electrophorus electricus* The organs, in one hour, hydrolyze amounts of ACh equivalent to one to five times their own weight. In the larger specimens, the organs have a weight of several kilograms, so that the

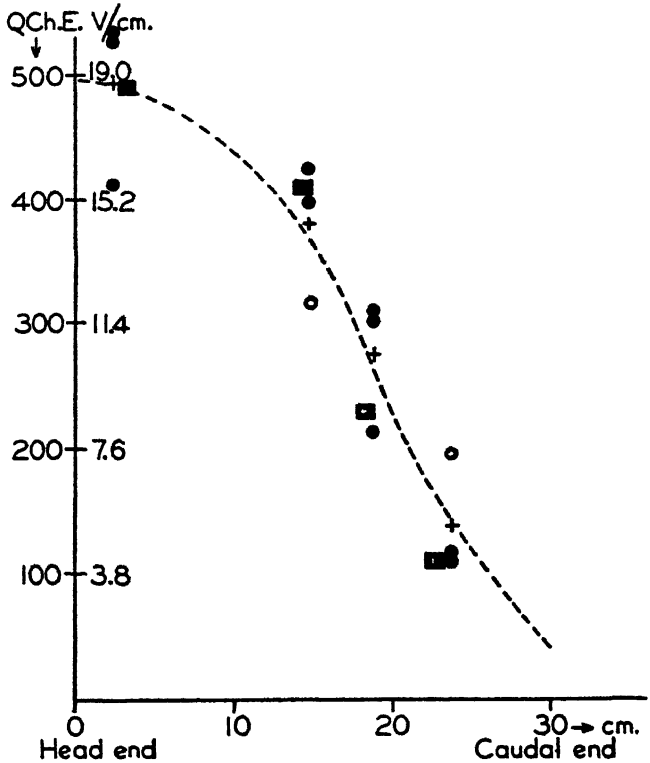


FIGURE 1. Action potential and cholinesterase activity in the electric organ, specimen no 1
Length of fish, 51 cm
Abscissae distance from the anterior end of the organ in cm
Ordinates: QCh E and V/cm.
● average QCh E from a single piece of tissue
+ average QCh E values from pieces of the same section
■ V/cm

amount of ACh which may be split in these organs may amount to several kilograms per hour or several milligrams in one-thousandth of a second. These are significant amounts. They make possible the assumption that ACh is directly connected with the action potential and may even generate it, for, in this case, the compound must appear and disappear in milliseconds. If speculation were to be excluded, the only means of removing this compound so rapidly would be by enzymatic action. The high concentration of a specific enzyme appeared partic-

ularly significant, in view of the chemical composition of these organs: They contain 92 per cent of water and only 2 per cent of protein.

In the weak electric organ of the common Ray, the concentration is relatively low. If, in the three species mentioned, voltage and number of plates per centimeter are compared with the concentration of cholinesterase, a close relationship becomes obvious.[15, 16]

A more detailed analysis has been carried out on the electric organ of *Electrophorus electricus*. This species is particularly favorable for such studies, since the number of plates per centimeter, and consequently, the voltage per centimeter, decrease from the head to the caudal end of the organ (PLATES 2 and 3). The cholinesterase activity

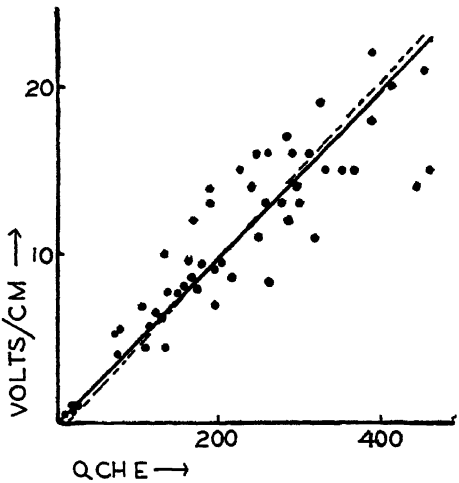


FIGURE 2 Correlation between voltage and cholinesterase activity.
The voltage per cm is plotted against the enzyme concentration. The dotted line is calculated from the data obtained with the method of least squares; the fully drawn line calculated on the assumption that the line goes through the 0 point.

decreases in the same proportion. If the electric changes are recorded and compared with the enzyme activity of the same section, a close parallelism is obtained between voltage and enzyme concentration (FIGURE 1). This is found not only in regard to the variations which occur in the same specimen, but even for the variations between the individuals, which are quite considerable.[17]

A great number of experiments have been carried out on fish of various sizes, covering a range of the action potential from 0.5 to 22.0 volts per centimeter. The quotient CH.E./V was found to be 20.7, with a standard deviation of only ± 0.7 or 3.7 per cent. The standard deviation for a single measurement is ± 5.1 or about 25 per cent. This is good uniformity for a quotient correlating physical and

chemical data. Of particular importance is the fact that the line correlating the two variables apparently goes through 0 (FIGURE 2). This indicates a direct proportionality. The results are consistent with the concept that the physical and chemical processes recorded are directly associated and, consequently, interdependent.[18] Such a parallelism has not been found with other compounds or enzyme activities known to be connected with nervous action.

The direct proportionality found between physical and chemical events is significant, in view of the changing morphological structure of the electric unit, the electric plate. If all plates were identical in structure, as *e.g.*, in the case of the electric organ of *Torpedo*, the voltage and the cholinesterase activity would be expected to be directly proportional to the number of plates. The situation is entirely different in the electric organ of *Electrophorus electricus*, because the structure of the plates shows enormous variations. In spite of all variations of the visible structure, the voltage of each plate is the same, namely, close to 100 millivolts It has, therefore, to be assumed that the "active membrane," with which the electric manifestations are connected and which is not yet well defined, does not change, but is similar in all plates. The direct proportionality found between voltage and enzyme activity suggests, then, that the physical and chemical events may be associated with the same membrane and that they may be functionally interdependent Here again, the fact is important only in connection with the great number of other observations, especially the extraordinarily high speed of the chemical process, without which the correlation observed would not have the same interest.

Two assumptions appear possible concerning the manner in which ACh may act: It may produce electromotive force directly by action on the surface, or it may decrease the resistance by increasing the permeability of the boundary. Resistance and electromotive force are closely related properties. So far, the evidence from experiments on nerves is in favor of a change in resistance and increased permeability. On the basis of alternating current impedance measurements carried out on the giant axon of squid, Cole and Curtis calculated that the resistance drops during the passage of the impulse from 1,000 ohms to about 25 ohms per square centimeter.[19] In experiments on the electric tissue, a comparable drop in resistance was found by Cox, Coates, and Brown.[20] There is no conclusive evidence that electromotive force is actually produced during the passage of the impulse. One possible interpretation on the basis of the material available at present is, therefore, the assumption that the parallelism found between voltage and

ACh metabolism may be due essentially to the effect of the ester on the resistance of the boundary or, which is equivalent, on its permeability.

Thus, we arrive at the following picture of the role which ACh may have in the mechanism of nerve activity: According to the membrane theory which is most widely accepted among physiologists, the nerve is surrounded by a polarized membrane. The polarized state of the membrane is due to a selective permeability to potassium ions which are many times more concentrated inside the axon than outside. During the passage of the impulse, the permeability of the membrane to negative ions is increased, and a depolarization occurs. The rapid appearance and removal of ACh may be an event essential for this change in permeability. The depolarized point becomes negative to the adjacent region, and flow of current results. This flow of current stimulates the next following point. There again, ACh is released, and the whole process repeated. The impulse is thus propagated along the axon. Cholinesterase destroys the active ester very rapidly, and the state of polarization may hereby be restored.

At the nerve ending, other factors, like increased surface and decreased resistance leading to a greater flow of current, may act in addition. But the process is fundamentally identical, the transmitting agent being the flow of current. Whereas, in earlier theories, ACh was considered as a "neurohumoral" or "synaptic" transmitter, *i.e.*, a substance released from the nerve ending and acting directly on a second neuron, in the new concept it is assumed that the transmitting agent is always the electric current, the action potential, but the release of ACh is necessary for generating the current.

The picture is consistent with the idea of the propagation of the nerve impulse as developed by Keith Lucas and Adrian. It becomes unnecessary to assume that the transmission along the axon differs fundamentally from that across the synapse. The assumption of a special mechanism at the synapse, different from that in the axon, as emphasized before, was the chief difficulty which had to be overcome to reconcile the original theory with the conclusions of the electrophysiologists This appeared necessary for any satisfactory answer to the problem. If it is true that physical methods alone are unable to explain the mechanism in a living cell, it is equally true that conclusions based on chemical methods should not be in contradiction to those obtained with physical methods, in view of the much higher sensitivity of the latter.

The picture of the transmission of the nerve impulse across the synapse is, however, far from being complete, if only the flow of current

from the nerve ending to the second unit is considered. The observations of Eccles and his associates have shown that the electric current set up by the pre-synaptic impulse initiates in the post-synaptic membrane a special junctional potential (end-plate potential or, more generally, synaptic potential[21]).

These findings have recently found a morphological correlate by the discovery of Couteaux that the sarcoplasm surrounding the presynaptic nerve ending has a very peculiar structure.[22] It is similar to that described by several authors of the last century in the electroplasm which surrounds the nerve endings in the electric plates of electric fish and which shows a layer of "electric rods," the "palisades" of Remak, at that particular point.

The biochemical data support the assumption of a high rate of ACh metabolism in the post-synaptic membrane of the neuro-muscular junction. At the motor end-plate of guinea pig gastrocnemius, only one third of the cholinesterase was found to disappear within three to four weeks after section of the motor nerve.[23, 24] The rest remained there for many months, a long time after the end-plate had been transformed into a sole plate. It appears, thus, probable that part of the high concentration of cholinesterase observed at the motor end-plate may be located at the post-synaptic membrane. The observations on the electric organ support the assumption of such a localization. The electric plates which form the electric organ are homologous to motor end-plates. The discharge in these organs can be considered as comparable to the end-plate potential, that is, a response of the post-synaptic membrane. The direct proportionality found between the voltage of the discharge and the cholinesterase activity is, therefore, another indication for the importance of ACh in the post-synaptic membrane

Specificity of Cholinesterase

In all the experiments on the activity of the enzyme, it was assumed that cholinesterase is specific for ACh. In such a case, not only is the conclusion justified that the substrate metabolized is ACh, but also, the activity of a specific enzyme determined *in vitro* may well be used as an indication for the potential rate of metabolism of the substrate occurring *in vivo*.

It appeared imperative, therefore, to demonstrate the specificity of the enzyme for ACh in all those tissues which were used in the investigations leading to the new concept. The ester linkage in ACh shows no peculiar properties. It has, therefore, to be expected that the ester can be hydrolyzed by other esterases and, on the other hand, that

cholinesterase can hydrolyze other esters Specificity, in this case, would be expected on the basis of analogy to be only relative, not absolute: Cholinesterase might be expected to split ACh at a higher rate than other esters, whereas other esterases might be expected to behave differently. By testing a number of substrates, a pattern has been obtained which makes it possible to distinguish specific cholinesterase from other esterases [25]

In the variety of nerve tissues which have been used as basis for establishing the new concept, the enzyme was found to be an esterase specific for ACh: *viz.*, mammalian brain, lobster nerve, squid fiber con-

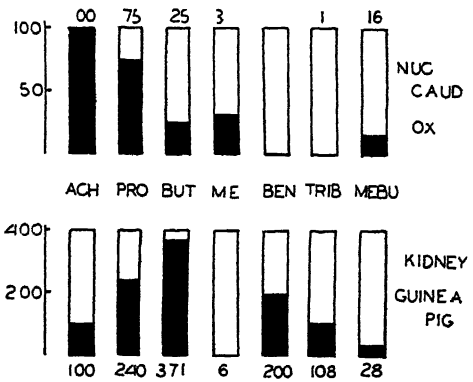


FIGURE 3 Pattern of cholinesterase (nucleus caudatus of ox) in presence of different substrates compared to that of an esterase (kidney) not specific for acetylcholine

The columns represent the Q of the substrates, the Q of ACh being 100 Abbreviations: Pr = propionylcholine, Bu = butyrylcholine, Me = acetyl-β-methylcholine (mecholyl), Be = benzoylcholine, Tr = tributyrin, Mb = methyl butyrate

taining the giant axon, and the electric tissue. All show a similar pattern, typical for cholinesterase. Even then, rigid statements should be avoided. Occasional deviations in one or the other directions may be expected. Recent observations of Richards and Cutcomp[26] have revealed that the cholinesterase of bee brain splits acetyl-β-methylcholine at a higher rate than ACh, whereas, otherwise, the pattern was typical for cholinesterase. In contrast, the hydrolysis patterns of the esterase of other organs (liver, kidney, and pancreas) differ greatly from that of cholinesterase (FIGURE 3). The esterase in these tissues shows several variations, but this could be expected, since the physiological substrate is unknown, and probably varies in the different organs. They should be referred to as *unspecified*, not as *unspecific*, esterases, because they may well be specific for substrates not as yet specified. ACh is metabolized at a high rate only in nerve tissue,

since only there is choline acetylase found If the esterase in all nerve tissue shows a pattern so distinctly different from that of the esterases of other tissues, it is justifiable to consider this enzyme as a specific cholinesterase.

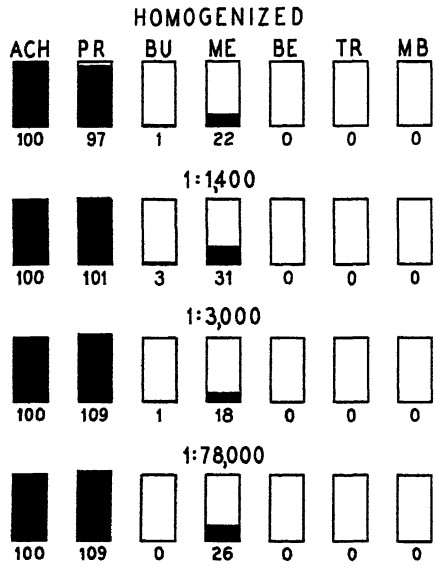


FIGURE 4 Rate of hydrolysis of different esters by the cholinesterase of the electric organ of *Electrophorus electricus*

The first row gives the data obtained with a homogenized suspension of electric tissue In such suspensions, 1 mg of protein splits about 20–40 mgs of ACh per hour
The three following rows show the data obtained with increasing degrees of purity, 1 mg of protein splitting 1,400, 3,000, and 78,000 of ACh per hour, respectively

Of particular importance is the question of the enzyme present in the electric tissue. The interpretation given for the direct proportionality between voltage and enzyme activity is justified only if the enzyme is exclusively, or almost exclusively, specific cholinesterase. Only in that case can the proportionality be referred to as an interdependence between ACh metabolism and electric manifestations.

The enzyme extracted from the electric organ of *Electrophorus electricus* has been purified by fractional ammonium sulfate precipitation. A high degree of purity may be obtained in this way. 1 milligram of protein is capable of splitting twenty to thirty thousand milligrams of ACh per hour. By further separation of the proteins by high speed centrifugation (ultracentrifuge), in collaboration with Dr. K. G. Stern, a degree of purity has been obtained where 1 milligram of protein was able to split eighty thousand milligrams of ACh per hour. If the rates

of hydrolysis of different substrates by the highly purified enzyme are compared to those obtained with the homogenized suspension of electric tissue, the pattern obtained remains exactly the same throughout the whole process of purification (FIGURE 4). Both have the pattern characteristic for cholinesterase. Thus, the correlation established between voltage and enzyme activity can be consequently referred to a correlation between ACh metabolized and voltage developed.

It may be noted that examination in the analytical ultracentrifuge indicates that the enzyme is a very large molecule. These are not yet final observations. If they could be confirmed, they would indicate that the turnover number of the enzyme is many millions per minute and that one molecule of cholinesterase could split one molecule of ACh within a few microseconds.

B. The Energy Source of the Nerve Action Potential

The second line of investigations, in which enzyme activity could be correlated with events in the living cell recorded by physical methods, is based on the energy transformations involved and on thermodynamic considerations.

If the release and removal of ACh are associated with the primary alterations of the nerve membrane during the passage of the impulse, then the primary source of the chemical energy released during the recovery process should be used for the resynthesis of ACh.

The most readily available source of energy in living cells is that released by energy-rich phosphate bonds. Phosphocreatine, the main "storehouse" of energy-rich phosphate bonds in muscle, is also present in nerves. The electric organ offers a suitable material for investigating the chemical reactions which supply the energy for the action potential. Both electric and chemical energy released are within the range of measurement, whereas, in ordinary nerves, such an analysis is difficult.

Measurements carried out on the electric organ of *Electrophorus electricus* have revealed that the chemical energy released by the breakdown of phosphocreatine is adequate to account for the electric energy released by the action potential. The electric energy released externally per gram and impulse, in large eels, was found to be 4 micro-calories. This is the maximum external energy which may be obtained, under the condition that the external resistance is approximately equal to the internal. The total electric energy is about 6 times as high as the external, or about 25 micro-calories. These data were obtained on eels of 170 to 180 cm length. In medium-sized eels of 90 to 120

cm. length, the total electric energy released per gram and impulse was found to be 47 micro-calories, on the average. There are some assumptions, on which these figures are based, which will be discussed by Drs. Cox, Coates, and Brown. If we consider all probable assumptions, these figures may possibly be revised downward by 15 per cent or upward up to 100 per cent.

Tested under the same conditions, the energy released by the breakdown of phosphocreatine was found to be 32 micro-calories per gram and impulse, in the large eels (average of 15 experiments). In the medium-sized eels, the energy released by phosphocreatine was about 51 micro-calories (average of 15 experiments). The lactic acid formation released about 17 μcal., in the large, and 53 μcal. in the medium-sized, eels per gram and impulse. The energy of the lactic acid is probably used to rephosphorylate creatine, just as in muscle where the phosphopyruvic acid transfers its phosphate *via* adenosine triphosphate to creatine ("Parnas reaction"). The sum of the two reactions may, therefore, be used as indication for the energy released by phosphate bonds. It amounts to 49 μcal. in the large, and 104 μcal. in the medium-sized, eels. The figures are consistent with the conclusion that energy-rich phosphate bonds are adequate to account for the energy of the action potential.

It appeared crucial to test whether or not energy-rich phosphate bonds are really the energy source of ACh formation. If this be the case, it would show that the energy of the primary recovery process is really used for the resynthesis of the compound which, by its release, supposedly initiates the nerve impulse. It would, therefore, at the same time, constitute a new support for the assumption that the primary "excitatory disturbance" which produces a propagated impulse may, indeed, be the release of the ester.

In confirmation of this assumption, a new enzyme, choline acetylase, could be extracted from brain in cell free solution, which, under strictly anaerobic conditions, in presence of adenosine triphosphate, forms ACh.[27-30]

The enzyme has been extracted from homogenized brain. From one gram of fresh rat or guinea pig brain, enzyme solutions were obtained which form 150–200 μg. of ACh per hour. More recently, up to 250 μg./g./hr. were obtained.

Presence of eserine and fluoride is necessary to inhibit the action of cholinesterase and adenosine triphosphate, respectively. Inhibition of the latter enzyme is necessary, since. otherwise, the breakdown of adenosine triphosphate occurs too rapidly. Fluoride inhibits this

breakdown, but it does not interfere with the transfer of energy-rich phosphate bonds, as has been shown by Ochoa.[31]

The enzyme has also been extracted from powder of acetone dried brain.[29, 30] Extracts prepared from one gram of powder form 1.0–2.0 mgs. of ACh per hour. Since acetone inactivates cholinesterase, this enzyme is largely, or sometimes completely, inactivated in the extracts prepared from powder of acetone dried brain, so that addition of eserine may have either a small effect or practically none on the formation of ACh. Adenosine triphosphate is also removed in extracts from acetone dried brain. No addition of fluoride is, therefore, required. For instance: In one experiment, 820 μg. of ACh were formed per gram and hour, with no eserine 780 μg, and without fluoride 810 μg. It has, thus, been demonstrated that the enzyme mechanism responsible for the formation of the ester is not identical with the hydrolyzing enzyme.

The enzyme requires the presence of potassium in high concentration, close to that found in brain. It contains active sulfhydryl groups which are readily inactivated by monoiodoacetic acid or copper in low concentration. The — SH groups are easily oxidized by air. On dialysis, the enzyme rapidly loses its activity. Addition of potassium ion and 1 (+) glutamic acid or cysteine reactivates partly. 1 (+) alanine, also, has some effect; other amino acids have either a weak effect or none. Citric acid has an effect nearly as strong as glutamic acid, whereas dicarboxylic acids have practically no effect.[28, 30]

The longer the dialysis is carried on, the weaker is the reactivation by the compounds mentioned. The experiments suggest that choline acetylase requires a coenzyme for its activity. The coenzyme has now been found. In contrast to the enzyme which occurs only in nerve tissue, the coenzyme has been extracted from brain, liver, heart, and skeletal muscle (Nachmansohn and Berman[32]). The coenzyme has been purified to a certain degree by treatment with barium salt, which precipitates the coenzyme. The purification, however, is still in progress. The coenzyme not only reactivates the dialyzed enzyme, but increases considerably the undialyzed enzyme preparations. Marked activation has been obtained in this way, especially in extracts from lobster nerve, rabbit's optic nerve, and electric tissue. Of special interest is the evidence for the presence of choline acetylase in the optic nerve. The possibility of a role of ACh in sensory nerves has been a matter of controversy for many years, since the ester was not found in such nerves, whereas cholinesterase is present in concentrations in an order of magnitude similar to that in motor nerves. The presence of choline acetylase in the optic nerve is further support for

the assumption that ACh may have the same function there as in other nerves.

The oxidation products of amino acids, *i.e.*, α-keto acids, have a strong inhibitory effect on the formation of ACh, when present in concentrations of 10^{-3} to 10^{-4} M. So far, pyruvic, phenylpyruvic, oxyphenyl pyruvic acid, and α-keto glutaric acid have been tested.[29, 30]

ACh formation has also been studied in extracts prepared from peripheral nerve fibers, in order to determine whether or not choline acetylase is present in the peripheral fibers, as well as in brain.[33, 34] This should be the case, if the new concept of the role of the ester in the axon is correct. It has been found that choline acetylase may be extracted from peripheral nerve fibers, as well as from brain. The rate of formation of ACh in extracts prepared from the sciatic nerve of the rabbit was found to be 70 to 90 μg. per gram and hour. The sciatic contains a large amount of inactive tissue (connective tissue, fat, and myelin). On the assumption that this tissue forms about two-thirds of the total weight, which is a conservative estimate, the amount of ACh which can be formed in the axon of the rabbit sciatic may, thus, be about 250 μg. per gram per hour, and is probably higher.

It appeared of special interest to determine the activity of choline acetylase during degeneration, and to test how this metabolism is related to the nerve function, *i.e.*, to conductivity. Conduction is still maintained two days after section, whereas, after three days, it has disappeared. If the release of ACh is responsible for conductivity, formation of ACh should be possible at a rate not too far below normal, as long as the nerve is capable of conducting.

Forty-eight hours after the section of the sciatic, choline acetylase activity has decreased only about 20 to 25 per cent. After seventy-two hours, when conductivity has disappeared, the decrease is marked, but still about one-third of the enzyme is present. The results are consistent with the assumption that enzyme mechanism is required for conduction.

C. Nerve Action Potential and Inhibition of Cholinesterase

In a third line of investigation, cholinesterase activity and nerve action potential could be directly correlated in experiments on the peripheral axon. One of the essential facts in support of the theory of "neuro-humoral" or "synaptic" transmission was the observation that ACh, when applied to synaptic regions, may have a stimulating action. No action has yet been obtained with the ester, when applied to the axon. Lorente de Nó[35] kept bullfrogs' sciatic nerve in a two

gram per cent solution of ACh for many hours, and did not find any effect on conductivity. He considers his failure to obtain an effect on the axon by ACh as proof against the new concept of the role of ACh in the mechanism of nerve activity. ACh is a quaternary ammonium salt. Such compounds are completely ionized and usually lipoid insoluble. Generally, they do not penetrate the lipoid membrane. Therefore, these compounds can be expected to have no effect on the axon, since axons are always surrounded by a lipoid membrane, even though it may be rather thin.

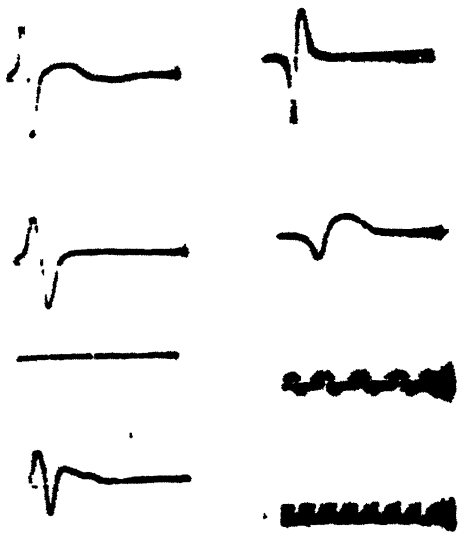


FIGURE 5. Effect of eserine on single fiber action potential (giant axon of squid)
Left: eserine 0.002M, records (from top to bottom) at 0′, 10′, 25′ (conduction abolished), 35′; sea water at 28′. Conduction distance less for last record, too short to demonstrate latency effect.
Right: eserine 0.01M, records at 0′, 15′, fiber then rinsed, and axoplasm analyzed chemically. Upper time scale applies to this experiment, lower to column at left, both 1000 c.p.s.

The problem has been approached in a different way: If ACh is the depolarizing agent and if the function of cholinesterase is to remove the active ester, so that polarization again becomes possible after the passage of the impulse, then inhibition of the enzyme should alter, and, in sufficiently high concentration, abolish, the nerve action potential.[36]

Eserine is known to be a strong inhibitor of cholinesterase. This compound is a tertiary amine and may, therefore, if undissociated, penetrate the lipoid membrane. Experiments carried out on the giant

axon and on the fin nerve of squid have shown that eserine alters, and, in higher concentrations, abolishes, the nerve action potential. Within a few minutes in eserine, amplitude, length, and duration of the action potential recorded with the cathode ray oscillograph are markedly changed, and in 20 to 25 minutes, the conductivity has been abolished (FIGURE 5). When the nerves are put back into sea water, they quickly recover, and conductivity reappears. The reversibility of the effect is consistent with the fact that the inhibition of cholinesterase is easily reversible *in vitro*.

Strychnine, another inhibitor of cholinesterase, was also found to alter, and, in higher concentrations, to abolish, the nerve action potential reversibly.

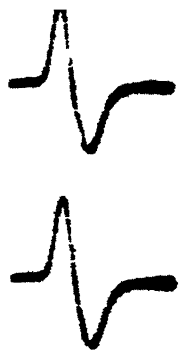


FIGURE 6. Effect of prostigmine on single fiber action potential (giant axon). Records before and after 45′ in 0.01M

Thus, a new relationship has been established between enzyme activity and nerve action potential, in this case using the peripheral axon.

Prostigmine has, *in vitro*, the same effect as eserine, but it has no effect on the nerve action potential (FIGURES 6 and 7). Prostigmine is like ACh, a quaternary ammonium salt, and it cannot penetrate the lipoid membrane. This has been demonstrated by the following experiment. The axoplasm of the nerves kept in eserine was extruded, and the presence of the compound was tested by the inhibitory effect on a purified cholinesterase solution. Even in thousand-fold dilution, the axoplasm from a portion of a single axon showed, by the inhibition

of esterase, easily detectable quantities of eserine. The axoplasm of nerves kept in prostigmine had no inhibitory effect on cholinesterase, even when undiluted.

Prostigmine, like ACh, has 3 methyl groups attached to the nitrogen. Drs. Bronk and Acheson have offered evidence that tetraethylammonium chloride acts on medullated nerve and, therefore, presumably enters it. This compound is also a quaternary ammonium salt and com-

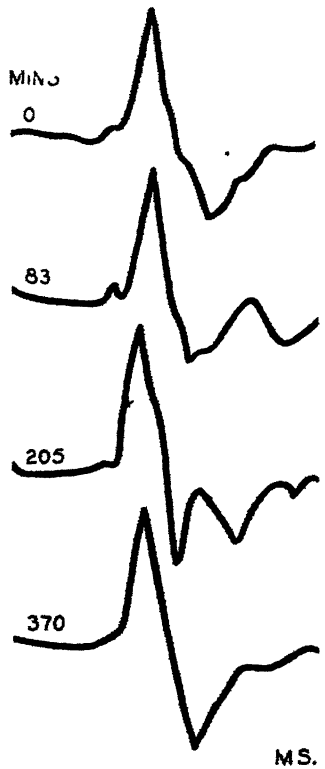


FIGURE 7. Effect of prostigmine on fin nerve Records before and after 83′, 205′, and 370′ in 0 01M Traced from enlarged photographs

pletely dissociated. Although it is true that ionized compounds are not readily soluble in lipoids, the properties which decide lipoid solubility are far from well established. Frequently, when, in a compound, the ratio of C over N is increased, it becomes more lipoid soluble. In tetraethylammonium chloride, there is four times as much carbon as in tetramethylammonium chloride. The change from

methyl to ethyl affects profoundly the physico-chemical properties of a molecule. The difference between methyl and ethyl alcohol is well known, and need not be discussed here.

Nearly half a century ago, Michaelis showed that, if, in a certain dye, the ethyl groups were substituted by methyl groups, no staining inside the living cell could be obtained.[37] Since these groups are not part of the molecule which has the staining properties, the loss of staining power may be due to the impossibility of penetrating the cell, due to the substitution performed. It appears not surprising that a compound with 4 ethyl groups becomes lipoid soluble, in spite of nearly complete dissociation.

The inability to penetrate the lipoid membrane may explain why ACh and prostigmine, applied externally, act only on nerve endings which do not have a myelin sheath, but are inactive when applied to the axon. Only in electric tissue may the power of ACh to produce an action potential be demonstrated. Injection of ACh leads to changes in potential of the same direction as those observed during the discharge.[38] Electric tissue, however, is an accumulation of end-plates which, in contrast to the axons, are not protected by myeline and, therefore, do react. This may also be the explanation for the famous observation of Claude Bernard on the effect of curare, since, according to recent observations, the active principle of curare is a quaternary ammonium salt.[39, 40]

The peculiar ability of the synapse to react to injected ACh can no longer be referred to a difference in the fundamental physico-chemical process underlying the propagation of the nerve impulse, but to the difference in histological structure.

Effect of Di-Isopropyl Fluorophosphate (DFP)*

Recently, a new inhibitor of cholinesterase, di-isopropyl fluorophosphate (DFP), became known, which can inhibit cholinesterase irreversibly. Tested on the fin nerve of squid, the compound has the same effect on the action potential as was observed with eserine, and at about the same concentration.[41] When the nerve is kept in a solution of 2 mgs. of DFP per cc., the action potential is completely abolished in about 30 minutes. When the nerve is put back into sea water, the action potential comes back (FIGURE 8). These experiments suggested that, for relatively short periods and at low temperature, around 20° C., the inactivation of cholinesterase by DFP may be partly re-

* Most of the observations reported in this paragraph were carried out after the conference, but, since the effect of fluorophosphate on the action potential and its mechanism played an important role then, it appeared desirable to include these data in this paper.

versible. The cholinesterase activity in squid nerves, under the experimental conditions used, could not be determined, since the season was too advanced, and no squids were available. Experiments were therefore carried out with the abdominal nerve cord of the lobster. This nerve preparation has a high cholinesterase activity and relatively satisfactory action potentials. The potentials recorded were exclusively those of the giant axons of the cord. The transmission across the synapses in the ganglia does not, therefore, enter into the picture.



FIGURE 8. Effect of DFP on the action potential of the fin nerve of squid. DFP 0.013M. First two records (from top to bottom): before, and after, 35′ in DFP. The last two records: recovery after 60′ and 215′ in sea water.

When the nerve is immersed in a solution of 2 mgs. of DFP per cc., the action potential disappears within about 30 to 40 minutes, as in the case of squid nerve. If the nerve preparation is put back into sea water, the action potential reappears after some time. Nerves kept in DFP for additional periods after the abolition of the action potential show less complete recovery. Exposure of the nerve to DFP, for 90 minutes after the disappearance of the action potential, abolishes the response irreversibly.

Determinations of cholinesterase in these nerves reveal a striking parallelism between the recovery of the action potential and the reappearance of cholinesterase (FIGURE 9). The less complete the re-

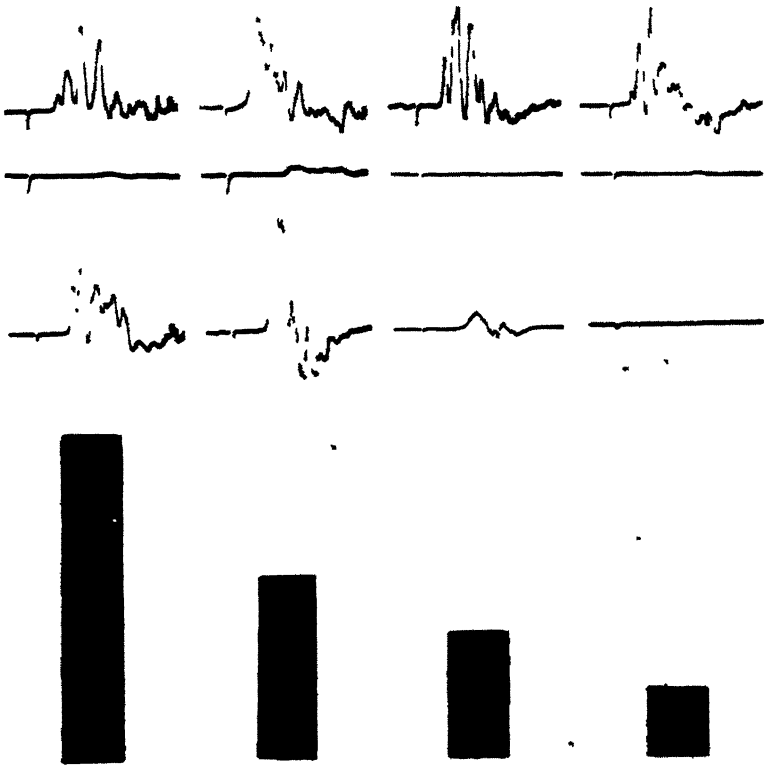


FIGURE 9. Reversibility of action potential and reappearance of cholinesterase in nerves exposed for varying periods of time to DFP, 0.013M.

The nerve whose action potentials are shown in Column 1 was transferred to sea water immediately after the action potential was abolished, and washed for one hour. The nerves of Column 2 to 4 were kept in DFP for 30′, 60′, and 90′, after the action potential had disappeared, and then washed in sea water. The top line of each column shows the action potential in the untreated nerves. The second line shows the abolition of the response by DFP. The third line shows the degree of recovery after washing the nerve. The reappearance of cholinesterase activity is shown in the vertical bars of the fourth line. The CO_2 output is 233, 129, 88.5, and 50 cmm. per 100 mgs. per hour.

covery of the action potential, the smaller is the amount of cholinesterase activity. Even after complete and irreversible abolition of the action potential, a small amount of enzyme activity may still be detected. The experiments indicate that cholinesterase inhibition by DFP of cold-blooded animals is partly reversible, for a certain period of time.

This has been confirmed by observations on *in vitro* inhibition of cholinesterase solution. DFP was added, in two different concentrations, 0.1 and 0.5 μg. per cc., to cholinesterase solution prepared from electric tissue. At the low concentration of DFP, the enzyme solution

liberated 520 cmm. CO_2, instead of 790 cmm. CO_2 per hour without DFP. The activity decreased slowly, over a period of hours. In the solution exposed to the stronger concentration of DFP, the activity was only 25 per cent and was nearly completely abolished after 30 minutes. If, after varying periods of exposure of the enzyme solution to DFP in greater concentration, the solution was diluted, part of the activity could be retained for a period of two to three hours (FIGURE 10). These experiments give additional evidence that the irreversible inhibition of cholinesterase by DFP is a slow process at low temperature.

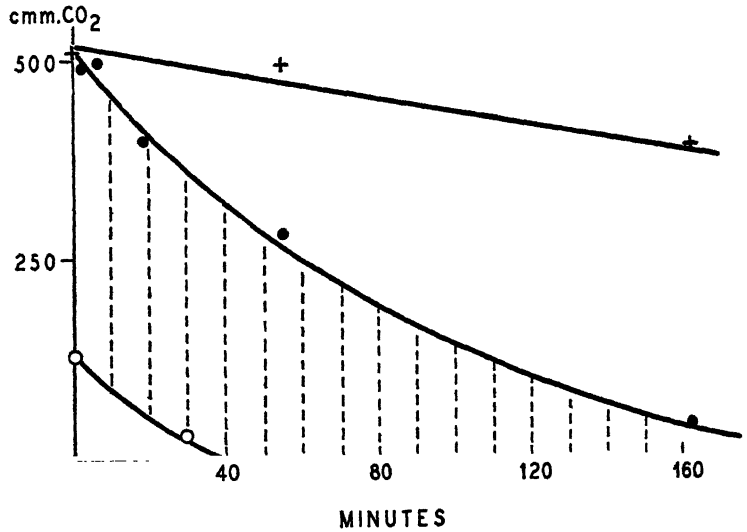


FIGURE 10. Reversibility of cholinesterase inhibition by DFP *in vitro*, tested by dilution effect, t = 9° C

The cholinesterase solution used liberates 790 cmm CO_2 per hour.
+ — + Activity found in presence of 0 1 μg of DFP per cc
O — O Activity found in presence of 0 5 μg of DFP per cc
● — ● Activity found after exposure to 0 5 μg of DFP per cc, for varying periods of time, and subsequent dilution to 0 1 μg per cc The part with the dotted lines indicates the reversibility as a function of time

Dr. Gilman presented observations on bullfrogs, in which it was found that, following injection of DFP, the action potential of the sciatic nerve may persist in the apparent absence of cholinesterase. The bullfrog sciatic nerve contains extremely small amounts of cholinesterase. 100 mgs. of nerve (wet weight) liberate 40–50 cmm. CO_2 per hour. Observations on lobster nerve indicate that the enzyme is present in about five times excess, since about 80 per cent may be removed while the action potential is unaffected. Even if, in the bullfrog sciatic nerve, the excess of enzyme is smaller when part of the

activity disappears, the measurement of the CO_2 liberation falls into a range where precise evaluation becomes difficult. Moreover, in such a preparation, the retention of CO_2 by the protein becomes an important factor. Finally, even in the thin lipoid membrane of the lobster nerve after prolonged washing, sufficient excess of DFP is retained to inhibit 20–40 per cent of the remaining esterase activity. At least this amount, if not more, may be retained in the relatively greater amount of myelin and fat in the bullfrog sciatic nerve. When this nerve is then ground, the retained DFP may come in contact with the cholinesterase and destroy a considerable fraction of the enzyme still present in the intact nerve.

DISCUSSION

It may be of interest to discuss the neuro-humoral theory in the light of recent developments, and to analyze the two basic experiments which form the main support for the hypothesis that the ester is actually liberated at the nerve ending and, having crossed the synapse or motor end-plates, acts directly on the second neuron or on the muscle fiber. The two observations are: (1) The stimulating action of ACh when applied to synaptic regions; (2) the appearance of the ester in the perfusion fluid, following nerve stimulation. It has just been explained why the effect of ACh applied externally is limited to the nerve ending. In any case, a stimulating effect is not necessarily a physiological effect, but may well be a pharmacological one. The same action may, indeed, be produced by other compounds. The observation of Otto Loewi that a compound appears in the perfusion fluid, following nerve stimulation, was important because it suggested that the compound may be connected with nerve activity. The importance of this observation need not be minimized because a quarter of a century later the original interpretation has to be changed. In fact, by the new development, the role of ACh became more general and more important than could originally have been anticipated. The appearance of a compound in the perfusion fluid, however, is not sufficient evidence for concluding that the compound acts outside the cell. Many compounds of intermediate cell metabolism may appear outside the cell. This is due to the fact that all enzymatic reactions follow a logarithmic curve. Therefore, if even the greatest part of a compound is rapidly metabolized by the intracellular enzymes, a small fraction may persist long enough to escape enzymatic action and leak out from the cell. This, apparently, may happen also to ACh, in spite of the high concentration of cholin-

esterase inside the cell, particularly when some kind of damage of the surface membrane is produced, as may be expected in the case of prolonged perfusion or in other unphysiological conditions affecting either the membrane permeability or the cholinesterase activity.

In order to verify the assumption that the amount of ACh actually released from the nerve ending is sufficiently high to produce a stimulating effect on the second unit, Dale and his associates attempted two sets of experiments. They determined the minimum required to produce a stimulus and compared it to the amounts released. However, in both cases tested, a puzzling discrepancy was found: In the case of the superior cervical ganglion, only 1/40,000 of the amount of ACh necessary to produce a single response appeared in the perfusion fluid per impulse. In the case of the muscle, only 1/100,000 of the amount of ACh necessary to produce a single twitch was collected. This difference is so considerable that the observations cannot be considered as evidence for the idea that ACh is the direct transmitter of the impulse, especially in view of all the other obstacles.

The situation is further complicated by the fact that these infinitely small amounts of ACh can only be found in presence of eserine which should inhibit their destruction. The enzyme located at the neuronal surface forms a barrier for the crossing of the ester. Even without regarding the existing discrepancy, it is very difficult to believe that, under physiological conditions, that is, in the absence of eserine, the small amounts of ACh released can cross the barrier and still arrive in sufficient concentrations for producing a response. The small amounts found under these conditions are easily explained if we assume that ACh is released inside the cell, and that the amounts which appear in the perfusion fluid are those which have escaped hydrolysis and have been preserved, due to the presence of eserine.

Another question on which some comments may be useful, is that of the difference between the rates of ACh formation and hydrolysis. There are two instances in which these two rates may be compared on the basis of experimentally established data: The guinea pig brain and the rabbit sciatic nerve. In the first case, about 200 to 250 μg. of ACh may be formed, whereas about 70 mgs. may be split per gram per hour. The rate of cholinesterase activity is, thus, about 300 to 350 times higher than that of choline acetylase. In rabbit sciatic nerve, the figures are about 100 μg. per gram per hour and 15–20 mgs. per gram per hour, *i.e.*, the rate of hydrolysis is about 150–200 times

as high as that of synthesis. It is doubtful whether these figures indicate the real difference of the possible rates of the two enzymes. Cholinesterase is an extremely stable enzyme. Its activity is determined in a well ground and homogenized suspension of the tissue. It appears probable that the maximal possible activity is actually measured *in vitro*. This is almost certainly not the case with choline acetylase. The enzyme is an extremely labile and a rather complex system which has to be extracted from the tissue. During the preparation, part of the activity may have been lost. We do not know whether the conditions used at present are optimal or even close to optimal. Although the enzyme was discovered three years ago, the rates of formation obtained are still continuously increasing, since more and more factors are becoming known which activate the enzyme (Nachmansohn and Berman, unpublished experiments). In such a case, it is possible and, in fact, probable, that the activity *in vivo* may be considerably higher than that observed in the solution. A sharp distinction has, moreover, to be made between the potential and the actual rate. Rates of enzymes measured *in vitro* are potential rates. The actual rates in the living cell may be entirely different. Many enzymes are present in excess in the cell. An excess of 3 to 5 times above the actual requirement is nothing unusual. A 5-fold excess of cholinesterase above that necessary for function has been recently observed in the case of lobster nerve (Bullock *et al.*[41]). Other enzymes are in much greater excess, whereas, in some cases, the excess activity appears to be relatively small. Nothing is known, at present, as to whether or not choline acetylase is present in excess. Even if this is the case, it may be much smaller than that of cholinesterase.

For an understanding of the problem, the decisive difference which has to be considered is not the difference of rates, but the difference of function. There is a fundamental difference between the function of cholinesterase and that of choline acetylase. If the release of ACh is an essential event in the alterations of the membrane during the passage of the impulse, then the active ester has to be destroyed within a millisecond or less, so that the resting condition may be restored. Therefore, the enzyme which removes the active ester, cholinesterase, has to be very active, but only during this brief period, and may then be inactive until the passage of the next impulse. The formation of ACh, on the other hand, need not be such a rapid process. It is generally assumed that the active ester is released from an inactive form. This is supported by the fact that the primary energy released during recovery is used for the synthesis of ACh, thus implying that the syn-

thesis is a slow recovery process. It is, therefore, not difficult to assume that the period required for the formation is longer than that for the hydrolysis of the same amount, according to the kind of nerve, its condition, temperature, and so on.

In the initial phase of nerve stimulation, the preformed ACh would act as a reserve and would make conductivity independent of the rate of ACh formation for a considerable length of time. Even the few μg. of ACh found per gram of nerve would be sufficient to make possible the passage of several thousand impulses. The actual amount of preformed ACh in the living cell may be higher than that found experimentally, since it is possible that, during the destruction of the cell, a process necessary for the determination, a considerable part of the preformed ester is destroyed. A nerve should, therefore, be able to respond to stimulation for a considerable length of time, independent of the rate of ACh formation. Only in cases of prolonged stimulation should the rate of formation become the limiting factor. If all preformed ACh has been exhausted, and stimuli are applied to mammalian nerve every five milliseconds, then the amount synthesized in the intervals between stimuli should be sufficient for producing the necessary alterations in the membrane when released by a stimulus, and should be equivalent to the amount actually destroyed, during the passage of the impulse, by cholinesterase. Since, in mammalian nerve, the duration of the spike is only 0.5 millisecond and the cholinesterase may have acted only during part of this period, *e.g.*, 0.1 or 0.2 milliseconds, a difference of 25 to 50 times between the actual rate of cholinesterase and that of choline acetylase activity would keep the nerve going indefinitely, if this were the only factor involved.

In summary, considering the difference between the rates of cholinesterase and choline acetylase, we have to keep in mind: (1) that there is a fundamental difference of function; (2) that the cholinesterase activity determined is probably the maximum possible, whereas the choline acetylase activity found *in vitro* is almost certainly below the optimal rate *in vivo;* (3) that the excess of cholinesterase may be greater than that of choline acetylase. In view of this situation, the difference between the rates found does not offer any difficulty and, in fact, appears close to that which one would expect of these two enzymes so different in function and properties.

As to the criticism of Dr. Gerard, who resolutely rejects the concept presented, some of his main objections may be discussed briefly. (1) The high speed required for any chemical reaction associated with the transmission of the nerve impulse has been considered for a long

time, by many leading physiologists as the chief difficulty for any chemical theory. It is gratifying to see that the evidence accumulated during the last ten years for the high rate of cholinesterase activity appears to be so impressive that Dr. Gerard now sees in this high speed one of the main difficulties. He calculates, for example, that the ACh preformed, plus that synthesized, could not possibly supply the ester as fast as cholinesterase can split it.

Such an objection would only hold if the whole amount of cholinesterase present were continuously and fully active. It appears likely, however, that, at any given moment, only part of the enzyme acts and only for extremely brief periods. The differences found between the rates of formation and removal of ACh appear, as pointed out before, to be well within the expected range.

(2) Still more puzzling to Dr. Gerard is the fact that, at the motor end-plate, there is 15,000 times more cholinesterase than in the surrounding muscle fiber, since there is no evidence for a great store or synthesis of ACh at this junction.

The difference between muscle fiber and end-plate is interesting, in view of the specialized localization. It is comparable to the distribution found in nerve between surface and axoplasm, which is infinite. In absolute amounts, the ACh which can be metabolized per impulse per end-plate is 0.000002 μg. The formation of this amount does not require a particularly powerful synthesizing system nor an intensive respiration. The energy required for the synthesis, even assuming a high frequency, would still amount to less than one per cent of the oxidative energy measured, a deviation which is far below the measurable range.

(3) Dr. Gerard assumes that the heat production by the ACh released would amount to 10 per cent of the total heat, whereas the initial heat is only about 3 per cent.

The frog's sciatic nerve is suitable for such a calculation, since here more experimental data are available than in other cases. According to von Muralt, 0.0006 μg. of ACh is released per gram per impulse.[42] This would amount to about 6×10^{-9} gram calories, which is 0.6 per cent of the total or 20 per cent of the initial heat.

(4) Finally, many other agents and enzymes are present in neurons, like adrenaline, thiamin, adenosine triphosphate, CO_2, and many others. Dr. Gerard asks how we can reasonably select ACh and assign to it alone an essential role in conduction of the nervous impulse. Undoubtedly, there are other compounds and enzymes playing an essential role in nerve activity. The ACh cycle is evidence for that. But none

of these compounds shows the typical features of the ACh system, like the high speed, the exclusive localization in the surface, the parallelism with voltage, etc. These unique features of the ACh system make it possible to associate the ester more closely with the action potential than all other agents so far known.

SUMMARY AND CONCLUSION

In view of the complex nature of biological mechanisms, one or two facts, however well established and suggestive, would not be sufficient for any theory. However, if a great number of facts point in the same direction, then they support each other and potentiate the value of each of them. The essential facts established may be summarized: (1) The high concentration of cholinesterase in nerve tissue makes possible the removal of ACh at a speed comparable to that of the electric manifestations. (2) Cholinesterase is localized everywhere at the neuronal surface where the bioelectrical phenomena occur. The exclusive localization in the surface contrasts strikingly with the localization of other enzymes. (3) Cholinesterase in nervous tissue (and in muscle) is distinctly different from all other tissue esterases occurring in the body. The enzyme, present in all types of nerves throughout the entire animal kingdom, shows similar properties. (4) A direct proportionality between voltage and cholinesterase activity has been established in the electric organ of *Electrophorus electricus*. (5) The primary energy source of recovery after the passage of the impulse, namely, the energy-rich phosphate bonds of adenosine triphosphate, is used for ACh synthesis. (6) The formation of ACh by choline acetylase occurs at a high rate in the peripheral fibers, as well as in the brain. The enzyme has, so far, been found exclusively in nerve tissue. (7) Anticholinesterases alter, and, in high concentrations, abolish, the nerve action potential. The abolition of the action potential is reversible, if the inhibition of cholinesterase is reversible; irreversible inhibition of cholinesterase abolishes the nerve action potential irreversibly.

These facts considered altogether make it highly probable that the release and removal of ACh is an intracellular event, directly associated with the nerve action potential.

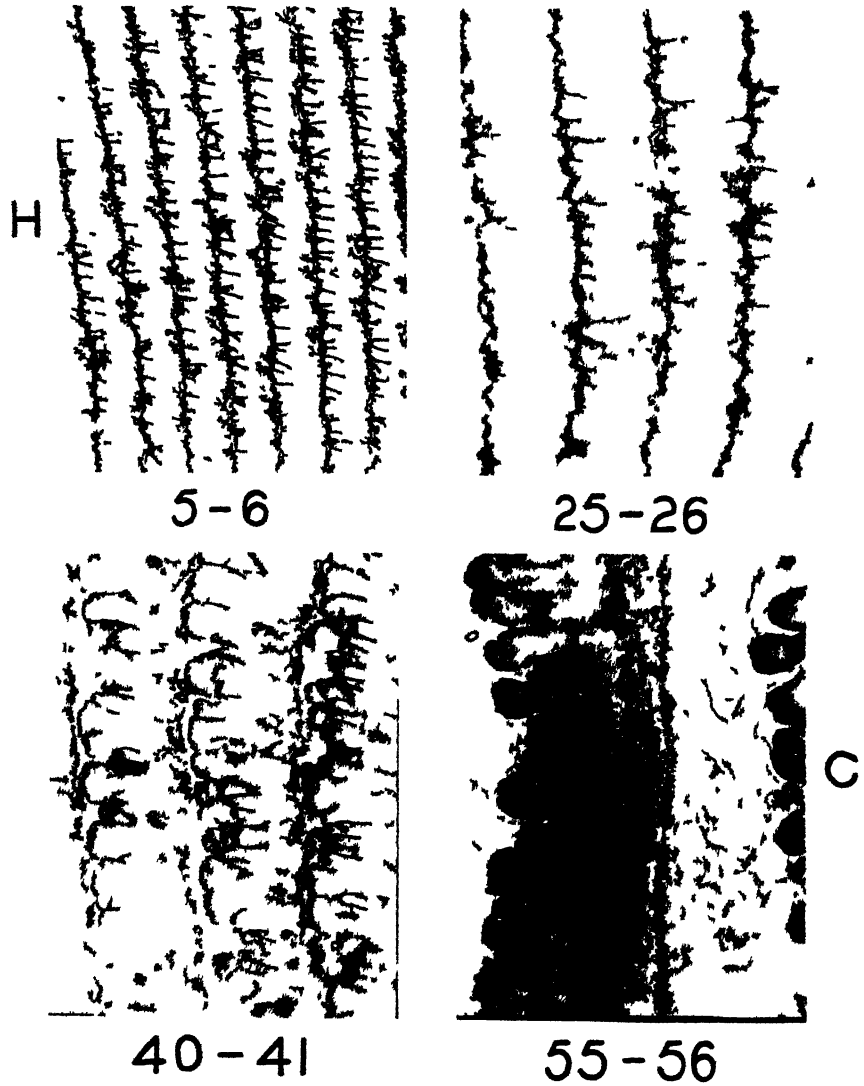
The precise function of the ester is still a matter of interpretation. On the basis of the physical and chemical data available, one possible interpretation appears to be that the ester plays an essential role in the breakdown of the membrane resistance, occurring during the passage of the impulse. New facts may change the situation. A number

of questions still have to be answered before a satisfactory picture of the chemical mechanisms of nervous action can be obtained.

BIBLIOGRAPHY

1. **Erlanger, J., & H. S. Gasser**
 1937. Electrical Signs of Nervous Activity. University of Pennsylvania Press. Philadelphia.
2. **Elliot, T. R.**
 1905. J. Physiol. **32**: 401.
3. **Loewi, O.**
 1932–1933. Harvey Lectures. **218.**
4. **Brown, G. L.**
 1937. Physiol. Rev. **17**: 485.
5. **Eccles, J. C.**
 1937. Physiol. Rev. **17**: 538
6. **Fulton, J. F.**
 1938, 1943. Physiology of the Nervous System. Oxford Univ. Press. New York.
7. **Nachmansohn, D.**
 1945. In: **Harris, R. S., & K. V. Thimann.** Vitamins and Hormones. Academic Press. New York. **3**: 337.
8. **Fulton, J. F., & D. Nachmansohn**
 1943. Science. **97**: 569.
9. **Nachmansohn, D.**
 1946. In: **Green, D. E.** Currents in Biochemical Research. Interscience Publishers. New York. 335.
10. **Marnay, A., & D. Nachmansohn**
 1938. J. Physiol. **92**: 37.
11. **Nachmansohn, D.**
 1940. Yale J. Biol. Med. **12**: 565.
12. **Boell, E. J., & D. Nachmansohn**
 1940. Science. **92**: 513.
13. **Nachmansohn, D., H. B. Steinbach, A. L. Machado, & S. Spiegelman**
 1943. J. Neurophysiol. **6**: 203.
14. **Hill, A. V.**
 1932. Chemical Wave Transmission in Nerve. Cambridge University Press. London.
15. **Nachmansohn, D.**
 1940. Science. **91**: 405.
16. **Nachmansohn, D., & B. Meyerhof**
 1941. J. Neurophysiol. **4**: 348.
17. **Nachmansohn, D., R. T. Cox, C. W. Coates, & A. L. Machado**
 1943. J. Neurophysiol. **6**: 203.
18. **Nachmansohn, D., C. W. Coates, & M. A. Rothenberg**
 1946. J. Biol. Chem. **163**: 39.
19. **Cole, K. C., & H. T. Curtis**
 1939. J. Gen. Physiol. **22**: 649.
20. **Cox, R. T., C. W. Coates, & M. V. Brown**
 1945. J. Gen. Physiol. **28**: 187.
21. **Eccles, J. C.**
 1943. J. Physiol. **101**: 465.
22. **Couteaux, R.**
 1944. C. R. Soc. biol. **138**: 976.
23. **Couteaux, R., & D. Nachmansohn**
 1940. Proc. Soc. Exp. Biol. & Med. **43**: 177.

24. Couteaux, R.
1942. Bull. biol. **76**: 14.
25. Nachmansohn, D., & M. A. Rothenberg
1945. J. Biol. Chem. **158**: 653.
26. Richards, A. G., & L. K. Cutcomp
1945. J. Cell. Comp. Physiol. **26**: 57.
27. Nachmansohn, D., & A. L. Machado
1943. J. Neurophysiol. **6**: 397.
28. Nachmansohn, D., H. M. John, & H. Waelsch
1943. J. Biol. Chem. **150**: 485.
29. Nachmansohn, D., & H. M. John
1944. Proc. Soc. Exp. Biol. & Med. **57**: 361.
30. Nachmansohn, D., & H. M. John
1945. J. Biol. Chem. **158**: 157.
31. Ochoa, J.
1941. J. Biol. Chem. **138**: 751.
32. Nachmansohn, D., & M. Berman
1946. J. Biol. Chem. **165**: 551
33. Nachmansohn, D., & H. M. John
1945. Science. **102**: 250.
34. Nachmansohn, D., H. M. John, & M. Berman
1946. J. Biol. Chem. **133**: 475.
35. Lorente de Nó, R.
1944. J. Cell. Comp. Physiol. **24**: 85.
36. Bullock, T. H., D. Nachmansohn, & M. A. Rothenberg
1946. J. Neurophysiol. **9**: 9.
37. Michaelis, L.
1900. Arch. mikr. Anat. **55**: 565.
38. Feldberg, W., A. Fessard, & D. Nachmansohn
1940. J. Physiol. **97**: 30.
39. King, H.
1935. J. Chem. Soc. London **2**: 1381.
40. Wintersteiner, O., & J. D. Dutcher
1943. Science **97**: 467.
41. Bullock, T. H., H. Grundfest, D. Nachmansohn, M. A Rothenberg, & K. Sterling
1946. J. Neurophysiol. **9**: 253.
42. von Muralt, A.
1945. Experientia **1**: 1.



NACHMANSOHN CHEMICAL MECHANISM OF NERVE ACTIVITY

PLATE 2

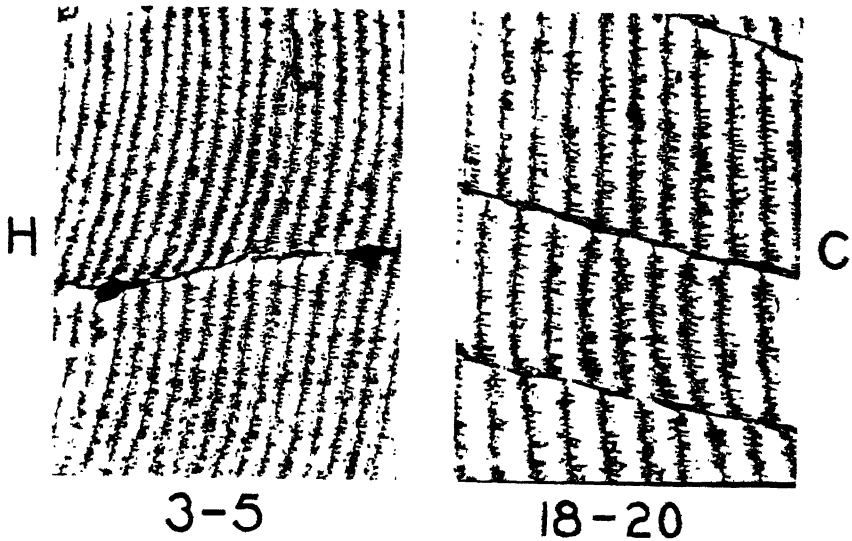
Changes of the morphological structure of the electric plates at different sections of the electric organ of *Electrophorus electricus*

The specimen used for this section was 114 cm long All sections are reproduced with the same magnification (× 145) The numbers below each section indicate the distance in cm from the anterior end of the organ H = head end C = caudal end

PLATE 3

Changes of the morphological structure of the electric plates at different sections of the electric organ of *Electrophorus electricus*.

The specimen used for this section was 57 cm. long (× 145). The numbers below each section indicate the distance in cm. from the anterior end of the organ. H = head end, C = caudal end.


H
C
3-5
18-20

AN ELECTRICAL HYPOTHESIS OF SYNAPTIC AND NEURO-MUSCULAR TRANSMISSION

By J. C. Eccles

University of Otago, Dunedin, New Zealand

1. PRESENT THEORETICAL POSITION

This paper will be restricted to the synapses of ganglia and the spinal cord and to the neuro-muscular junctions of skeletal muscle (henceforth, collectively referred to as synapses), because, physiologically, they form a fairly homogeneous group. A preliminary report has already been published.[24] There is now good evidence that the transmission of impulses, at all these synapses, is mediated by catelectrotonic potentials set up at the synaptic membrane of the post-synaptic cell—the end-plate potentials of skeletal muscle[26, 28, 34, 35, 50, 77] and the synaptic potentials of ganglion cells[22, 23] and motoneurones.[2, 3a, 25] We may, therefore, subdivide the problem of synaptic transmission into two Problems: (a) the mechanism whereby impulses in pre-synaptic nerve fibers set up catelectrotonic synaptic potentials in the post-synaptic cell; and (b) the initiation of impulses in the post-synaptic cell by these synaptic potentials. As is well known, the existing hypotheses relating to Problem (a) are chemical (acetylcholine), or electrical, or some combination thereof.[4, 70, 68] Problem (b) has, hitherto, been regarded as just a part of the general problem of impulse initiation by catelectrotonus. However, there is evidence of a unique mechanism in the case of the only synapse worked on in detail.[50]

There is some resemblance between these two stages of synaptic transmission and the two "boundary faces" postulated by Buchthal and Lindhard,[8, 9] to explain the two stages of neuro-muscular block produced by curare and acetylcholine.

In their existing form, both hypotheses relating to Problem (a) are unsatisfactory:

(i) Originally, the acetylcholine hypothesis simply stated that a pre-synaptic impulse liberated at the synapse a sudden jet of acetylcholine, which excited the post-synaptic cell by acting on specific receptors;[11] thus set up the synaptic potential, according to present views; and was quickly removed by the locally concentrated cholinesterase.[6, 7, 19, 67] The usual failure to detect acetylcholine in venous blood collected from

eserinized, stimulated ganglia or muscle suggested the additional hypothesis that, normally, acetylcholine is removed by being rebuilt rapidly to a precursor and that cholinesterase merely acts as a barrier, to prevent diffusion of acetylcholine away from the synapse.[62] Further additions to this hypothesis were needed, in order to explain the effects of eserine and curarine on end-plate potentials.* It is unsatisfactory that the acetylcholine hypothesis has had to be reconciled with new experimental evidence, by thus making subsidiary *ad hoc* hypotheses, which have not been independently testable. The most recent development of the acetylcholine hypothesis[68] is essentially a special type of the electrical hypothesis, for it postulates that electrical transmission across the synapse excites the *post-synaptic* liberation of acetylcholine, which, in turn, sets up the synaptic potential.

(ii) Most expressions of the electrical hypothesis of synaptic transmission have merely stated that the electrical currents of the pre-synaptic impulses set up impulses in the post-synaptic cell, much as one segment of a nerve excites the next.[20, 32, 60, 65, 66] The attempt at precise formulation by Lapicque (the isochronism hypothesis) has had to be modified so much by the recognition of the significance of additional factors (relative durations of the pre- and post-synaptic responses,[65] the rheobase of the post-synaptic cell[66]), that it now states little more than the above vague formulation. Thus, the electrical hypothesis is unsatisfactory (indeed, virtually useless), because it is so vaguely expressed that it fails to give predictions that would be a fertile source of experimental tests.

2. RECENT EXPERIMENTAL SUPPORT FOR AN ELECTRICAL HYPOTHESIS

The need for a more developed electrical hypothesis has now become urgent, because the following recent investigations have indicated that acetylcholine plays but a subsidiary role at ganglionic synapses, and a negligible role at spinal cord synapses. With muscle, too, there are indications that electrical transmission may play an important part.

(i) A detailed study[4, 24] of the electrical responses of eserinized ganglia (normal and curarized) revealed an excitatory action, prolonged for several seconds after repetitive stimulation and attributable to acetylcholine. However, this prolonged action was so weak that the summation of about 20 volleys in quick succession was needed to make it sufficiently strong to excite normal, fully eserinized ganglion

* Cf. Eccles, J. C., B. Katz, & S. W. Kuffler.[27]: 227–8

cells to discharge. The usual transmission mechanism was due to an excitatory action, unaffected by eserine and brief enough to be attributable to the action currents of the pre-ganglionic impulses.[22, 23]

(ii) A similar study of synaptic potentials of motoneurones excited through mono-synaptic reflex pathways of the spinal cord (frog, cat) has failed to detect even such a subsidiary role for acetylcholine transmission.[23] Furthermore, it has been found that synaptic transmission of the frog's spinal cord is unaffected by prolonged soaking (several hours) in high concentrations of acetylcholine (up to 1 in 5,000). Still higher concentrations have an anesthetic action which, initially, is reversible. The isolated oxygenated cord (anesthetized or unanesthetized) is soaked for 30 min. in a strong anti-cholinesterase (1 in 50,000 eserine), and then the acetylcholine is added to the solution.[31] These experiments would appear to falsify the hypothesis that acetylcholine plays a major role in synaptic transmission in the spinal cord. However, too much emphasis should not be placed on these latter experiments, until they are repeated with prostigmine as an anti-cholinesterase (cf. iii, below).

(iii) Just as with sympathetic ganglia,[23] the responses of curarized, eserinized (or prostigminized) muscles to repetitive stimulation are sharply distinguishable into a prolonged end-plate potential which is attributable to acetylcholine, and an initial, very brief, end-plate potential, but little lengthened by anti-cholinesterases.[27, 36] It seems probable that, as with ganglia, the small, apparent lengthening of the initial phase by anti-cholinesterases may be attributable to some admixture of the prolonged acetylcholine phase, and that the initial phase may be excited by the action currents of pre-synaptic impulses (cf. PART 8, ii). Acetylcholine blocks neuro-muscular transmission,[5, 9] presumably by catelectrotonic blockage, but, despite a relatively high acetylcholine background (1 in 200,000), pre-synaptic volleys still set up large end-plate potentials, even larger than in curarized muscle.[36] When performing these experiments by soaking frog's sartorii in acetylcholine solutions, prostigmine is used as an anti-cholinesterase. Eserine is ineffective, probably because acetylcholine competes with it for the cholinesterase.[27*]

3. EXPERIMENTAL BASIS FOR ELECTRICAL HYPOTHESIS

In recent years, important advances have been made in the investigation of nerve and muscle fibers, and an electrical hypothesis of transmission must be based on the following evidence:

* Cf. Eccles, J. C., B. Katz, & S. W. Kuffler.[27]: 225.

A. The Electrical Properties of the Surface Membranes and the Changes Produced by Catelectrotonus, Anelectrotonus, Local Responses, and Propagated Impulses

Quantitative measurements have been made of resistance, electromotive force, capacity, and rectification. The great diminution of the two former during the excited phase of the impulse has been described for nerve and muscle, vertebrate and invertebrate.[3, 12, 13, 16, 17, 37, 46, 47] There is no good evidence that the large inductance of cephalopod nerve[12, 14] is present in normal vertebrate nerve or muscle. The phenomenon of para-resonance is simply explained in terms of the two excitation constants of nerve.* In contrast with cephalopod nerve, there is, in frog muscle, no appreciable lag between a sudden change in potential and the associated change in the resistance of the membrane.[47] Such a lag in cephalopod nerve has been attributed to the large membrane inductance.[12] On present evidence, vertebrate nerve and muscle may, therefore, be regarded as having a negligible inductance, and may be provisionally schematized, as in FIGURE 1.

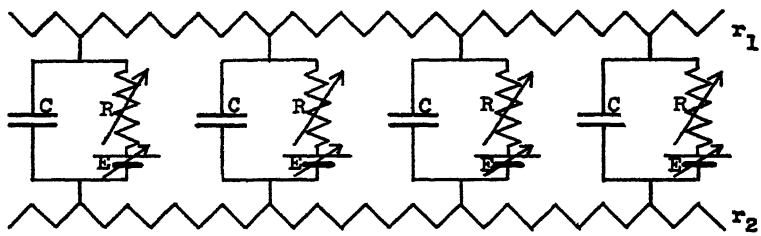


FIGURE 1. Diagram showing probable electrical characteristics of nerve and muscle membranes, r_1 and r_2 being, respectively, the external and internal longitudinal resistances; C, E, and R, the capacity, battery, and resistance of the membrane.

B. Local Responses

All grades of active local responses, short of propagated, all-or-nothing impulses, have been shown to exist in nerve and muscle,[3, 37, 42, 44, 46, 47, 52, 70, 79] and may be explained as due to the limited area excited[76] and or to the low intensity of the excitation.[12] In refractory, anesthetized, or deteriorated nerve or muscle, these local responses may be very large.[42, 47, 63, 70, 79] It appears probable that active local responses differ from passive electrotonic changes (including rectification), just as with the propagated impulse, in that they are caused by a temporary diminution, extinction, or even reversal, of the membrane battery.[3, 15, 16, 76]

* Katz, B.[46]: 28.

C. Reactions of Ephapses (Artificial Synapses)

The double axon preparations[1, 48, 19] are particularly relevant to electrical action across synapses. Since there has been excellent correlation between the effects predicted by the "local current" theory of nerve conduction and the effects observed, it may be concluded that these effects are caused by electrical current flow across the ephapse.[48, 49, 64] According to the geometry of the ephaptic contact, three main types of effect are exerted on the resting fiber by an impulse in the active fiber:[1]

(i) *At regions where fibers are contiguous for some distance on either side—for maximum effect, at least half a wave length.* Here the currents generated by the impulse have, in turn, anodal, cathodal, and anodal action on the resting fiber.[48, 64] FIGURE 2 shows that the

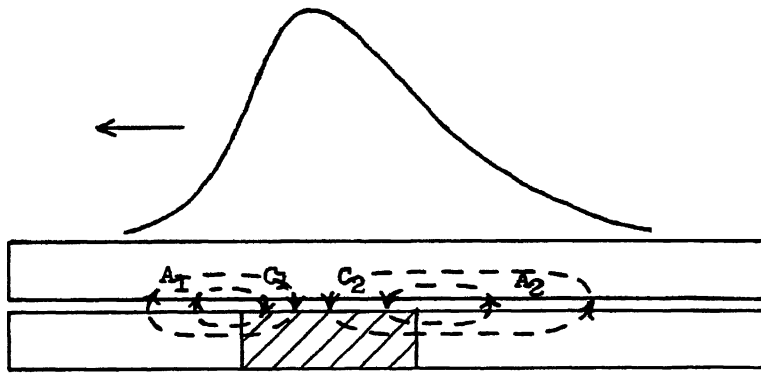


FIGURE 2. Diagram of two contiguous fibers, showing the current flow generated by impulse in active lower fiber and its penetration of the resting fiber (cf Katz & Schmitt, FIGURE 2[48]) As the impulse (shown above) propagates along the active fiber, any point on the resting fiber is subjected, in turn, to effects $A_1C_1C_2A_2$. Active part of impulse shown by hatched area in this and subsequent figures.

cathodal phase is really double, being due to currents generated by the membrane battery of the active fiber, at first ahead (C_1), and then in the wake, of the impulse (C_2). Thus, the sequence of action is $A_1C_1C_2A_2$, as the impulse sweeps past a point on the resting fiber. As Katz and Schmitt pointed out, the current penetrating and acting on the resting fiber is virtually a mirror image of the penetrating current of the active fiber, and hence, has an intensity-time course corresponding to the second derivative of the monophasic action potential, with respect to the time coordinate (d^2P/dt^2). Since the curvature of the wave front of the impulse is at least twice as sharp as that of its wake, A_1 and C_1 will be at least twice as large as C_2 and A_2, but the latter will be correspondingly longer in duration.*

* Cf. Katz, B., & O. H. Schmitt.[48] FIGURES 5A2 and 6, Curve 1.

(ii) *At regions where the resting fiber is not affected by the approach of the impulse, but only by its immediate juxtaposition and its propagation.* As shown in FIGURE 3a, this occurs when the impulse propagates from an electrically insulated region of the active fiber to a region where it is in contiguity with the resting fiber. Effects A_1 and C_2 are prevented by the insulation, the interaction being due to effects C_1 and

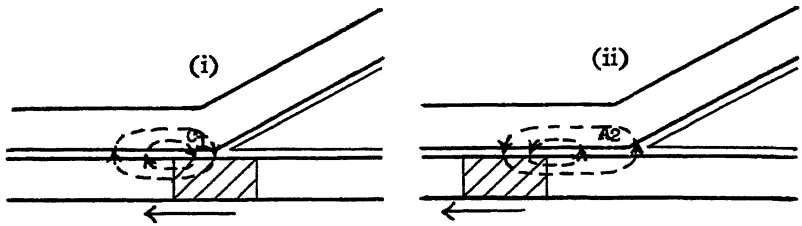


FIGURE 3a Diagram showing current flow at junctional zone of two previously separated fibers (i) Impulse at junction gives C_1 effect on resting fiber, (ii) after further propagation, wake of impulse gives A_2 effect

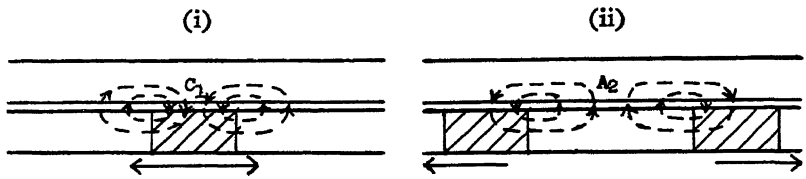


FIGURE 3b Penetrating current generated by impulse arising in one fiber (i) C_1 effect, as impulse is initiated; (ii) A_2 effect, in wake of impulse, propagating in both directions from site of origin

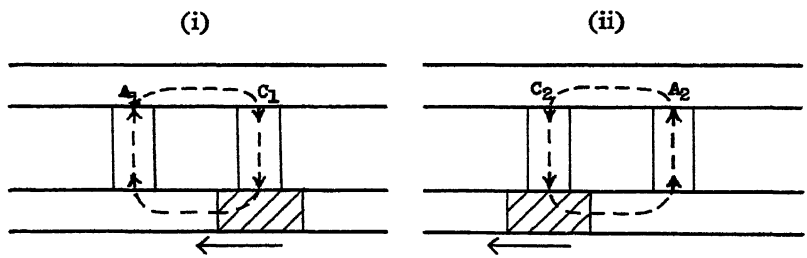


FIGURE 3c Two fibers connected by double salt bridge (i) Impulse opposite proximal arm, (ii) impulse opposite distal arm showing C_1A_2 action on resting fiber at proximal arm and A_1C_2 action at distal arm

A_2* A similar effect would also be produced at a region in a passive fiber adjacent to the origin of the impulse in the active fiber (FIGURE 3b), and at the proximal arm of a double salt bridge (FIGURE 3c).[69]

(iii) *At regions where the resting fiber is influenced by the approach and juxtaposition of an impulse, but not by its propagation.* FIGURE 4

* Cf. **Arvanitaki, A.**[1] FIGURE 4 III; **Marrazzi, A. S., & R. Lorente de Nó.**[64] FIGURE 3

(also FIGURE 3c, for distal arm of double salt bridge) shows that this occurs under the converse conditions to those above, and that effects

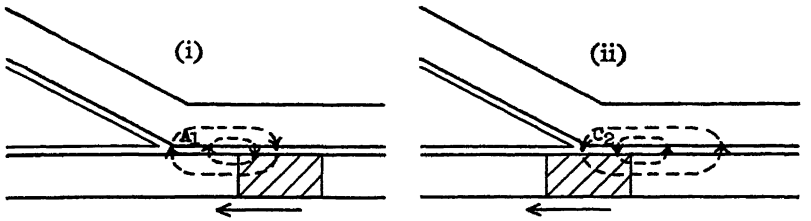


FIGURE 4a Converse diagram to FIGURE 3a, showing current flow at zone of separation of two previously contiguous fibers (i) Impulse approaching bifurcation gives A_1 effect, (ii) impulse at bifurcation gives C_2 effect.

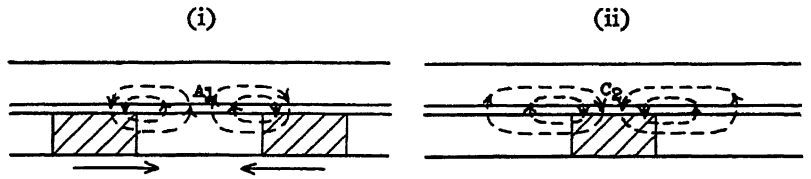


FIGURE 4b (i) Two impulses approaching a collision give A_1 effect on resting fiber, and (ii) at collision of impulses, C_2 effect on resting fiber

C_1 and A_2 are prevented, the action being due to effects A_1 and C_2.* At synaptic regions, a similar electrical action would be exerted by the pre-synaptic impulse on the post-synaptic membrane (cf. FIGURE 5), since the surrounding conducting medium provides a pathway for currents generated by the approaching impulse. There is, of course, no departing impulse (*boutons de passage* excepted). In this connection, Arvanitaki's results[1] are of especial interest, for it was only in ephaptic situations giving A_1C_2 effects that she observed appreciable local responses of the resting fiber She concluded that, in all other conditions, the terminal A_2 effect suppressed any active response of the resting fiber.

The approximate time-course of the penetrating current at a synapse may be derived from the monophasic spike potential at the pre-synaptic terminal, by considering the flow of current, much as Katz and Schmitt[48] did for two parallel fibers. The longitudinal current flowing in the external circuit is proportional to the first derivative of the monophasic potential, dP/dt. Immediately proximal to the end of the pre-synaptic fiber, all the longitudinal current is provided by the current penetrating the terminal end. Hence, this penetrating current is also proportional to dP/dt,† and not to d^2P/dt^2, as occurs along the length of the fiber. A

* Cf. **Arvanitaki, A.**[1] FIGURE 4 II.
† Cf. **Marrazzi, A. S., & R. Lorente de Nó**[66]: 89.

similar time-course may be assumed for the current penetrating the closely adjacent post-synaptic membrane (see FIGURE 5). The first derivative of the monophasic potential gives, of course, the expected diphasic effect, A_1C_2.

Thus, it may be concluded that, so far as they go, ephaptic investigations lend support to the hypothesis that an excitatory action would be exerted by impulses terminating at synapses. However, with ephapses, this excitatory action is normally too weak to initiate impulses in the resting fiber. For example, Katz and Schmitt[48] find that the maximum C effect is never as much as 20% of threshold, and Arvanitaki[1] has to sensitize the resting fiber by decalcification, in order to increase the local response sufficiently for impulse initiation.*

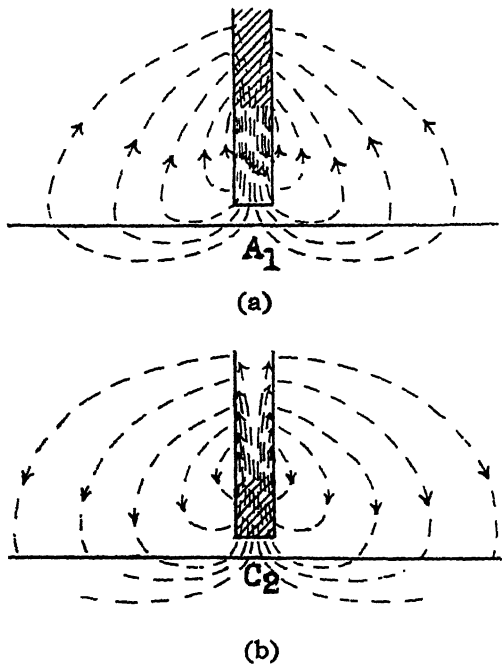


FIGURE 5. Diagrams of current flow at a schematic synapse with pre-synaptic impulse approaching synapse in (a), and at synapse in (b). Note reversal of current flow, the focal A_1 effect being followed by the focal C_2 effect at the synaptic region of the post-synaptic membrane.

The Hering effect, ephaptic transmission adjacent to killed or injured regions,[38, 43, 73] may also be explained as due to A_1C_2 stimulation of fibers rendered sensitive by the catelectrotonus[74] prevailing close to the injured region. Tests of excitability changes, 4 mm. from the killed

* Cf. **Arvanitaki, A.**[1] FIGURE 5.

end, showed a terminal A_2 effect,[64] but, presumably, this would disappear, closer to the killed end. The ephaptic experiments, in general, show that special conditions must prevail at synaptic contacts, if electrical excitation is to be adequate for synaptic transmission (cf. PART 8).

D. Special Properties of the Synaptic Region

So far, such investigation has been restricted to the isolated neuromuscular junction. When electrical recording is effectively localized to the end-plate region of the muscle, it has been shown that the end-plate potential set up by a nerve impulse rises smoothly to the full height of the spike potential,[50] without showing the sudden inflection characteristic of impulse initiation.[42, 32, 79] The impulse appears to be initiated, a little later, by an adjacent region of the membrane, when it reaches a critical intensity of catelectrotonus. Progressive curarization progressively diminishes the end-plate potential; the impulse initiation occurs adjacently, after the longer delay ensuing before the lower end-plate potential builds up the critical catelectrotonus; and eventually, transmission fails. It may, therefore, be assumed that the end-plate region of the muscle is specialized to give "local responses" of high and graduated intensities, without the sudden incursion of the all-or-nothing "breakdown" of resistance and battery that occurs with impulse initiation.[50] This evidence of unique electrical properties of the end-plate is relatable to its well-known, unique, pharmacological properties.[53, 54, 55] In the isolated preparation, Kuffler[53] has failed to detect the large resting potential (positive or negative) between the surface of the end-plate and that of the muscle fiber that has been described by Buchthal and Lindhard.[8]

4. INITIAL ASSUMPTIONS OF HYPOTHESIS

The following three initial assumptions of the electrical hypothesis are based on the evidence of the preceding four sections, together with the conventional histological picture (they form, as it were, a model of a synapse whose functional operation will be discussed in PART 5):

A. That the geometrical situation at the synapse may be schematically represented by the pre-synaptic fiber ending as a cylindrical membrane, with a closed end in close apposition to the large plane surface membrane of the post-synaptic cell, as is shown in section in FIGURE 5. Histologists are now fairly generally agreed that a transverse membrane exists at the synapse,[10] and there is also electrical evidence* of a

* Eccles, J. C.[30]: 352.

highly resistant transverse membrane. This evidence for ganglionic synapses also obtains for neuro-muscular junctions.

B. That, in general, the surface membranes of FIGURE 5 have the electrical properties demonstrated for peripheral nerve and muscle membranes: resistance, electromotive force, capacity, and rectification, as shown in FIGURE 1. There are no direct observations on nerve cells, but they resemble nerve fibers in their electrical excitability and in the propagation of impulses from a nerve cell to its axon,[20, 58, 60] and vice versa.[20, 56, 60, 72] It may also be assumed that both the exterior and the interior of the cells are good conducting media, and that the resting potential of the post-synaptic membrane is identical with that for the remainder of the post-synaptic cell.[53]

C. That the synaptic region of the post-synaptic cell has unique electrical properties, in that cathodal polarization (lowering of resting charge) sets up a graduated "local response," with a temporarily irreversible and large diminution of electromotive force and resistance, but not the all-or-nothing membrane "breakdown" characteristic of the propagated impulse (cf. PART 3, B and D). Direct evidence is only available for the end-plate region,[30] but the assumption is extended to the synaptic regions of nerve cells.

5. DEVELOPMENT OF HYPOTHESIS ON THE BASIS OF THIS MODEL

It appears that, assuming A and B, we have to expect that the current generated by an impulse propagating up to the terminal of the presynaptic fiber will, in part, penetrate the post-synaptic cell and give a diphasic action (cf. PART 3, C, iii). Firstly, there will be an anodal focus, A_1, at the synaptic region, with a cathodal surround (FIGURES 5a and 6a). Then, when the active region of the impulse reaches the terminal, current flow will reverse, giving a cathodal focus, C_2, with an anodal surround (FIGURES 5b and 6b). The penetrating current will be limited by polarization of the membrane and, in the initial phase, by the increasing resistance of the localized anode (rectification effect). On account of its much larger area, the membrane resistance (and penetrating current density) at the cathodal surround will be so much lower that its simultaneous diminution by the rectification effect will be relatively insignificant in tending to increase the flow of penetrating current. However, in the second phase, the situation is reversed, because, on account of the high current density, the lowering of the initially high resistance at the localized cathode will have a preponderant effect in increasing the flow of current. Hence, due to rectification, the

over-all resistance offered to the penetrating current will be much lower in the second phase than in the first, the current being, as it were, canalized through the localized low resistance at the cathodal focus. It should be noted that, in this way, rectification will diminish, at the synapse, the depressing action of the relatively high intensity A_1, and then increase the stimulating action of the relatively low intensity C_2 (cf. FIGURE 7b). The effectiveness of this discriminative action of rectification is illustrated for bipolar stimulation, by Cole.* It should be even more effective for the unipolar type of stimulation that occurs at the synapse. It is evident that, if the membrane had a high inductance in series with R (FIGURE 1), the brief penetrating currents would be much less intense, and the rectification correspondingly less effective.

The polarization of the membrane, in the first and second stages, is shown diagrammatically in FIGURES 6a and 6b. Note the wider spread of anelectrotonus, A_1, than C_2, and the reversal of potential gradients along the inner side of the membrane, corresponding to the reversal of the "core currents" (cf. FIGURES 5a and 5b). Note, also, that, at the dotted lines separating the anelectrotonic and catelectrotonic areas, the curves of the inner and outer membrane potentials are inflected, as would be expected for zero density of penetrating current. The catelectrotonic focus shown in FIGURE 6b will not immediately develop the pre-synaptic current flow reverses. The anodal polarization in FIGURE 6a takes some time to be removed by the local current flow, as well as by the reversed penetrating currents, and further time is needed to charge the membrane condensers to the fully-developed cathodal focus in FIGURE 6b (cf. FIGURE 7b). If, at this latter stage, the external electric field, applied by the impulse in the pre-synaptic terminal, were suddenly removed, the membrane would immediately revert to the potential distribution of FIGURE 6c (assuming that the internal and external media have equal longitudinal resistances; *i.e.*, that r_1 and r_2 of FIGURE 1 are equal). If no local response is set up, *i.e.*, if the membrane exhibits only its electrotonic properties, local current flow would quickly cause the anodal surround to discharge into, and repolarize, the cathodal focus, and the membrane would quickly revert to the normal, uniformly charged, condition. Thus, under such circumstances, with the usual disposition of electrodes for recording responses at the synaptic region (one close to the synapse and one distally on the post-synaptic cell), there would be recorded merely a brief diphasic potential, attributable to currents generated by the pre-synaptic impulse and but little modified by the passive properties of the post-synaptic cell.

* Cole, K. S.[13] FIGURE 3.

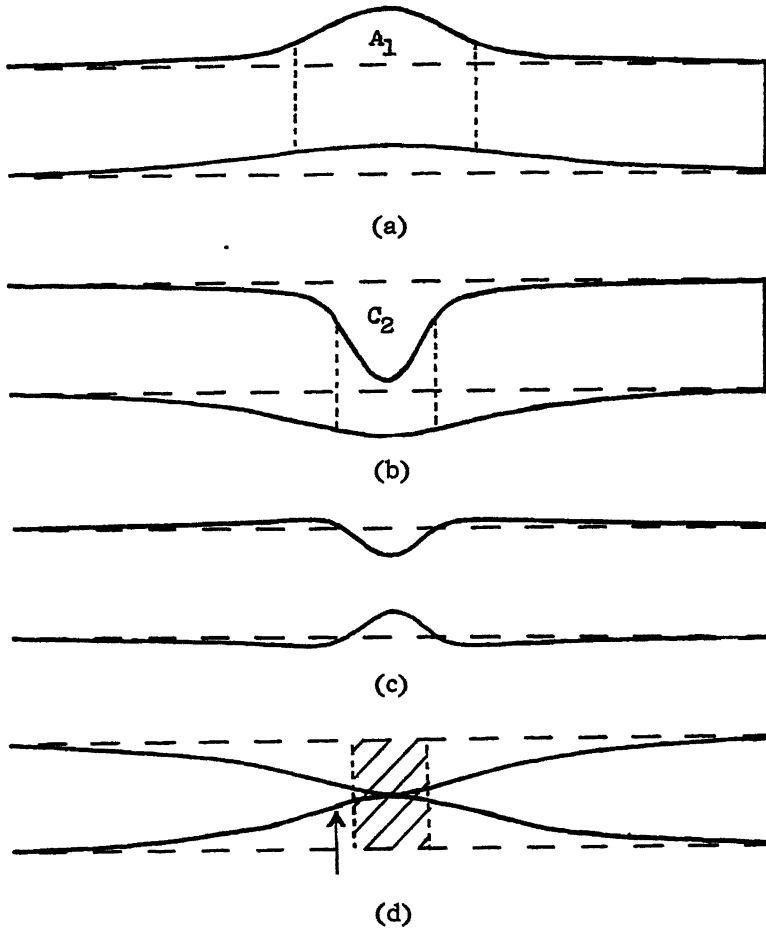


FIGURE 6 Graphs of spatial distribution of potentials on the outer and inner sides of post-synaptic membrane, the synaptic region being in center, i.e., potentials are ordinates and distances abscissae. The reference potential is given by a distal region of the membrane, the outer side being shown above the inner. The normal resting condition is shown by the broken lines separated by the resting potential (a) Initial A_1 focus at synapse with low intensity cathodal surround. (b) Reversed phase with C_2 synaptic focus and anodal surround (cf FIGURE 5b). (c) Membrane potentials, when external field generated by pre-synaptic impulse is removed, the spatial distribution of the potentials *across* the membrane being identical with those of (b). (d) Potentials after generation of local response at synapse, with the catelectrotonus (the synaptic potential) spreading thence over the post-synaptic membrane Hatched area shows specialized synaptic zone of membrane Impulse initiation occurs outside this zone, for example at the arrow.

There would be no synaptic potential, with its characteristic long duration.

The additional assumption (PART 4, C) is necessary, in order to explain the origin of the synaptic potential. It postulates that, when above a critical intensity, the cathodal focus evokes at the synaptic

region of the post-synaptic cell an intense "local response," which runs through a cycle of increasing and decreasing intensity, much as does the spike of a propagating impulse.[42, 52, 63, 70, 79] This local response, thus, would outlast the second phase of the penetrating current-flow, and provide a relatively enduring focus, of very low polarization (presumably due to diminution or extinction of the membrane battery) and resistance, through which adjacent regions of the post-synaptic membrane proceed to discharge. Such a local response is actually observed at an ephapse giving A_1C_2 interaction.[1] FIGURE 6d shows the internal and external potentials of a fully-developed local response at the end-plate region, where Kuffler[50] finds the potential as high as the spike potential. It is shown as zero transverse membrane potential, since it is not known if reversal of potential occurs with the muscle spike. Also, in FIGURE 6d, the anodal surround (of FIGURE 6c) has given place to a catelectrotonic surround of diminished polarization, which spreads spatially, according to "core conductor" theory. With muscle, it appears that the all-or-nothing spike arises when the membrane adjacent to the end-plate is critically depolarized,[50] *e.g.*, at the arrow (FIGURE 6d). The synapses of ganglion cells and of motoneurones of the spinal cord have also been observed to generate such catelectrotonic potentials (synaptic potentials), spreading spatially, according to core conductor theory.[2, 22, 25] It has further been shown that, as with the end-plate potential,[26, 77] these potentials have a time-course, which may be interpreted as due to a brief, active polarization and a passive exponential decay governed by the electric time constant (the product CR, in FIGURE 1) of the membrane.[22, 25] In the present hypothesis, the active depolarizing action is provided by the local response of the synaptic region of the post-synaptic cell, not directly by the currents generated by the pre-synaptic impulse, as has hitherto been assumed in hypotheses of electrical transmission.[20, 47, 65, 77, 78] There have, however. been suggestions of a possible involvement of a local response.*

The recent hypothesis of Nachmansohn[68] is relevant to assumption C, for it would postulate that the local response is due to the action of acetylcholine liberated by the post-synaptic membrane at a critical intensity of catelectrotonus (*i.e.*, at X in FIGURE 7c). Such an assumption is readily assimilable to the present electrical hypothesis, but its general application to nerve impulse transmission would seem to be falsified by Lorente de Nó's finding[61] that this transmission is unaffected by high concentrations of acetylcholine.

* Eccles, J. C.[21]: 369; Lorente de Nó, R.[60]: 449; Arvanitaki, A.[1]: 103.

The further problem of synaptic transmission concerns the initiation of propagated impulses by the catelectrotonic synaptic potential (see PART 1, Problem (b)). The observations of Kuffler[30] on the isolated neuro-muscular junction indicate that the initiation of the impulse occurs in the muscle membrane adjacent to the end-plate region (cf. FIGURE 6d). No such intimate observations have been made for synapses in ganglia and the spinal cord, but it may be assumed that the catelectrotonic potentials from diverse synaptic regions sum by electrotonic spread. So far, only the over-all, summed potential has been observed after electrotonic spread along the axon (the so-called synaptic potential.[2, 3a, 22, 23, 25]) In setting up the discharge of impulses, this synaptic potential appears to act just as a catelectrotonus, the discharge occurring at a critical degree of depolarization. The synaptic potential provides a satisfactory explanation of all the phenomena hitherto attributed to the central excitatory state.[2, 22,25] The "detonator response"[20, 21] need no longer be considered as a separate entity, for that hypothesis was based on experiments now explicable, in part, by the flow of penetrating current, as in FIGURE 5, and, in part, by the postulated local response of the post-synaptic membrane.[21, 22, 25]

Thus, the sequence of events in synaptic transmission is envisaged as:

(1) Impulse in pre-synaptic nerve fiber generates a current which gives a diphasic effect at the synaptic region of the post-synaptic cell, with a total duration of probably not more than 1 msec. in mammalian muscle and the spinal cord; initial anodal focus, with cathodal surround; more intense cathodal focus, with anodal surround.

(2) This cathodal focus sets up a brief and intense local response at the synaptic region.

(3) From this local response, a catelectrotonus spreads decrementally over the post-synaptic cell membrane.

(4) A propagated impulse is set up in the post-synaptic cell, if this catelectrotonus is above a critical value. If it is below, then, as the local response subsides, the catelectrotonic surround decays passively.

6. APPLICATIONS OF THE HYPOTHESIS

The hypothesis offers an explanation of the following observations on synaptic transmission:

A. Irreversibility of Synaptic Transmission

This may be explained in the following three ways: (i) An impulse, artificially set up in the post-synaptic cell and fired antidromically at

the synapse, would, in general, propagate past the synapse and so exert on the pre-synaptic nerve terminals the full sequence of $A_1C_1C_2A_2$, the terminal A_2 cutting short any excitation of the pre-synaptic terminal by C_1C_2;[1] (ii) the asymmetry of the pre- and post-synaptic elements, both as regards relative size and convergence relationship;[1] (iii) the pre-synaptic terminal may not have the special excitatory properties postulated for the synaptic region of the post-synaptic cell (assumption C). Of these, (i) is susceptible to test with the neuro-muscular junction. Antidromic transmission across synapses has been observed only under the special conditions provided by the prolonged end-plate negativity set up by nerve impulses in eserinized muscle.[27] There, the nerve terminal would be sensitized by the currents generated by the localized end-plate negativity.* Moreover, some of the muscle impulses may be blocked at the end-plate by the catelectrotonus.[33, 35] Thus, such antidromic transmission occurs under conditions resembling those causing ephaptic transmission, close to a killed or injured region of nerve (the Hering effect, PART 3, C, iii). The present hypothesis would predict that antidromic synaptic transmission would be greatly facilitated by colliding two muscle impulses at the end-plate region (cf. FIGURE 4b). The pre-synaptic fiber would then be subjected to the greatly increased excitatory action of double strength A_1C_2 stimulation, and antidromic transmission should occur under much less favorable predisposing conditions.

B. Synaptic Delay

If the initiation of the post-synaptic impulse is always caused by mediation of a synaptic potential of the post-synaptic cell, then true synaptic delay measures the interval between the time of arrival at the synapse of the fore-front of the pre-synaptic impulse and the initiation of the synaptic potential. In FIGURE 7a, the time-course of the action potential at the pre-synaptic terminal is shown, and below it (the dotted line in FIGURE 7b), the first derivative, which gives the approximate time-course of the current penetrating the post-synaptic membrane (PART 3, C, iii). Allowance for rectification action is made in the broken line of FIGURE 7b. On account of the electric time constant of this membrane, its potential change (the continuous line in FIGURE 7b) is shown lagging behind the current which produces it.† Now, according to the hypothesis, the post-synaptic membrane initiates a local response when the catelectrotonus reaches a critical value, *e.g.*, at the point X, in FIGURE 7c. As shown in FIGURE 6d, this

* Cf. **Eccles, J. C., & J. L. Malcolm.**[30] FIGURE 15a.
† Cf. **Katz, B., & O. H. Schmitt.**[41] FIGURE 6.

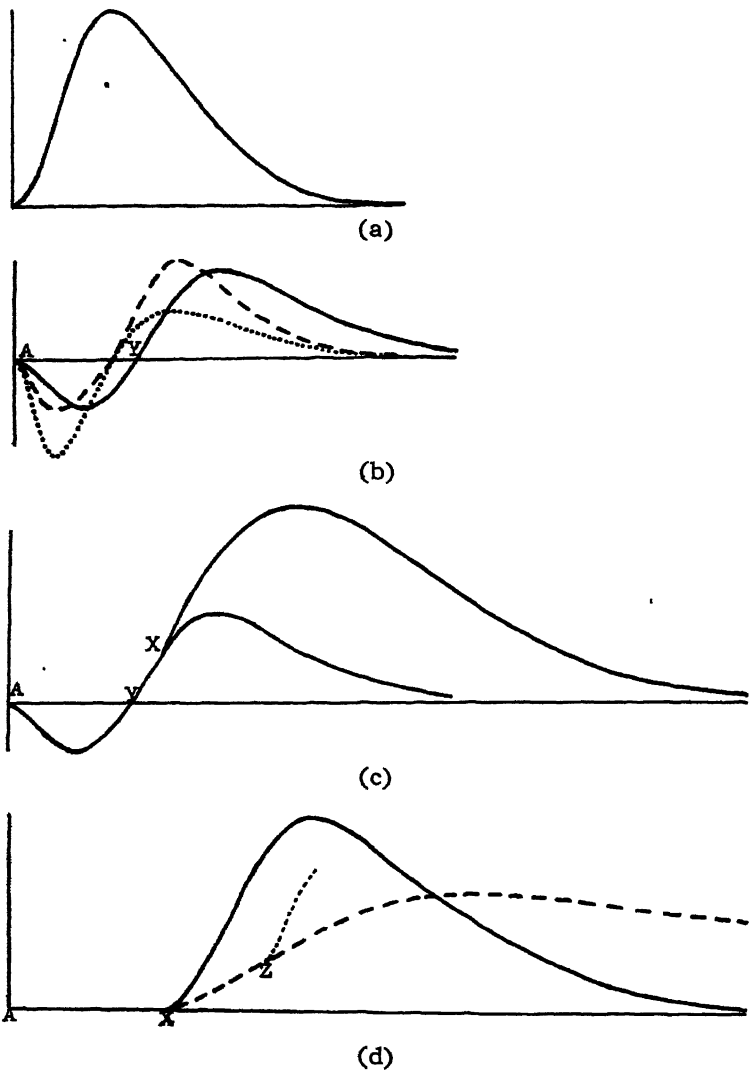


FIGURE 7. (a) Monophasic pre-synaptic action potential. (b) Its first derivative (dotted line) giving the time-course of the post-synaptic penetrating current (PART 3, C), anodal currents being plotted downwards. The broken line shows the modification produced in this current if resistance is doubled at the anodal focus and halved at the cathodal focus (rectification). The continuous line gives the approximate time-course of the post-synaptic membrane potential so produced, allowance being made for the electric time constant of the membrane (cf. Katz & Schmitt[48], FIGURE 6). (c) Post-synaptic membrane potential shown as in (b). At X, the local response of the synaptic membrane is initiated by the catelectrotonic phase, and it is shown running a time-course rather slower than a spike. (d) Neglecting the passive electrotonic changes of (b) and (c), this local response is plotted together with the time-course of the resulting spreading catelectrotonus, shown as the broken line (the synaptic potential) recorded from the post-synaptic membrane adjacent to the synapse (*e.g.*, at arrow, in FIGURE 6d). The rising phase of an initiated impulse is also shown (dotted line), AZ being its synaptic delay.

local response provides the active region for setting up a spreading catelectrotonus—the synaptic potential. Thus, the synaptic delay (in FIGURE 7, c and d), is represented by the interval AX. The value of 0.6 msec. which has been observed for this synaptic delay at mammalian neuro-muscular junctions[26, 28] and motoneurone synapses[25] (frog, 1.3 msec.) accords well with the duration of the pre-synaptic action potential (FIGURE 7a), when it is remembered that there is probably some slowing of time-course, as the impulse propagates into the fine pre-synaptic terminals.[59] As shown in FIGURE 7d, a further delay, XZ, usually 0.2 to 0.3 msec., is involved in the building up of the synaptic potential to the threshold, for initiating an impulse at Z.[25, 26, 28]

On the basis of FIGURE 7, the hypothesis offers a satisfactory explanation of all the experimental findings on synaptic delay. For example: (i) By facilitation, synaptic delay cannot be shortened below a limiting value of about 0.5 msec. for central synapses.[59] In FIGURE 7b and 7c, AY would be the minimal interval at which excitation could occur, under optimal conditions of facilitation. (ii) Synaptic delay can, however, be further shortened by the direct excitatory action of a preceding subliminal induction shock.[58] By its depolarizing action, the shock would diminish the time lag between reversal of current and reversal of potential, and so shorten synaptic delay to less than AY (FIGURE 7b). (iii) The longer synaptic delay with sympathetic ganglia (about five times longer) correlates with the longer duration of the pre-synaptic spike,[40] which sets the time scale throughout FIGURE 7. (iv) The upper limiting value of synaptic delay, for example, about 1.0 to 1.5 msec. for mammalian central synapses[25, 57, 59, 60, 71] and neuromuscular junctions,[26, 28] has been correlated with the time of the rising phase of the synaptic potential,[25, 26] and thus, according to FIGURE 7d, to the duration of the local response of the post-synaptic membrane (cf. C, below). (v) Synaptic delay (neuro-muscular in frog)[26] has, as would be expected from FIGURE 7, approximately the same temperature coefficient (2.1) as has the duration of the spike potential.

C. Time-Course of the Active Phase of the Synaptic Potential

The time-course of a local response is but little slower than the spike potential.[42, 52, 63, 70, 79] According to the hypothesis, therefore, the brief phase of active polarization (determined by analysis of the synaptic potential) should have a time-course somewhat slower than the spike of the post-synaptic cell. This accords well with the findings on ganglion cells,[22] motoneurones,[25] and muscles.[26] Furthermore, the temperature coefficient of this "active phase" is approximately the same as for a spike.[26]

D. The Brief Period of Low Resistance

This is in addition to that attributable to catelectrotonus, during the initial, "active" phase of the end-plate potential.[47] The postulated local response of the end-plate region of the muscle would produce just such an additional fall of resistance, running the same time-course as the active phase.*

E. Slow Catelectrotonic Potentials in the Pre-Synaptic Fibers[2, 30]

Such potentials have only been observed with synapses in the central nervous system (the dorsal root potentials). The present hypothesis has been extended to explain the production of these potentials, by making the additional assumption that the terminal region of the pre-synaptic fiber resembles the post-synaptic area, in being specialized to give local responses. The catelectrotonic focus provided by the local response of the post-synaptic membrane sets up the current, which excites the pre-synaptic terminal to give a local response, which, in turn, acts as a focus, setting up the spreading catelectrotonus of the dorsal root potential.[30]

7. DIFFICULTIES OF THE HYPOTHESIS

The electrical hypothesis encounters difficulties in explaining the following experimental observations, but possible lines of reconciliation are suggested.

A. Synaptic Block Produced by Curarine in Skeletal Muscle and Ganglia

Curarine acts as a specific depressant of the excitatory responses evoked in motor end-plates and ganglion cells by acetylcholine and related substances.[6, 7, 18, 27, 53, 54] Thus, the acetylcholine hypothesis provides an obvious explanation of the synaptic blockage produced by curarine. Now curarine causes such blockage by depressing the synaptic potential.[22, 26, 28, 50] Hence, according to the electrical hypothesis, the simplest explanation of the blockage would be, that there is depression of the local response set up by the cathodal focus (cf. FIGURES 6d and 7c); *i.e.*, that curarine depresses the electrical excitability of the post-synaptic membrane, as well as its acetylcholine excitability. In a recent attempt to test this, by electrical stimulation of the motor end-plate in the isolated nerve-muscle fiber preparation, before and after curarization, the initiation of propagated muscle im-

* Cf. Katz, B.[47] FIGURE 11.

pulses was used as the criterion of end-plate excitability. Yet, if the end-plate region reacts by local responses, rather than by propagating impulses (PART 4, C), it seems probable that this investigation tested the excitability, not of the end-plate (*i.e.*, of the post-synaptic membrane), but of the membrane adjacent thereto, confirming the previously observed absence, there, of curare action.[39, 75, 78] Thus, it provides no evidence for, or against, a specific depression of electrical excitability of the post-synaptic membrane. Alternatively, if curarine blocks *solely* by its known depressant action on the local potentials set up by acetylcholine,[18, 53, 54] then, in very deep curarization, the residual catelectrotonic effects produced by the cathodal focus, *i.e.*, by electrical transmission, should be observable, uncomplicated by chemical transmission. This inference is particularly pertinent in the case of sympathetic ganglia, where the synaptic potential is virtually abolished by deep curarization, and yet other evidence suggests that acetylcholine transmission plays but a minor role (PART 2, ii).[23]

B. Action of Anti-Cholinesterases on Synaptic Transmission with Skeletal Muscle and Sympathetic Ganglia

Anti-cholinesterases (eserine, prostigmine) delay the summit of the curarized end-plate potential and slow its decline,[27, 34, 35] effects which, undoubtedly, are attributable to a prolongation of the active depolarizing agent. With rapid, repetitive stimulation, a still greater effect is observed, the end-plate potential persisting for several seconds, both in curarized and normal muscle.[27, 46] With sympathetic ganglia (normal or curarized), anti-cholinesterases also cause a prolonged synaptic potential to appear, after rapid, repetitive stimulation, but this prolonged potential is sharply distinguishable as a special addition to the otherwise unaltered synaptic potential.[23] Presumably, both with ganglia and muscle, the prolonged potential is due to acetylcholine liberated by pre-synaptic impulses.* With ganglia, it has been argued that, since the initial, brief transmitter action, setting up the synaptic potential, is unaffected by anti-cholinesterases, it is not due to acetylcholine.[23] Similarly, with muscle, the eserinized (or prostigminized) end-plate potential appears to be the partly fused compound of a brief, initial phase, but little, if at all, lengthened by the anti-cholinesterase and the prolonged phase (certainly due to acetylcholine).† Thus, the electrical hypothesis would attribute the effect of anti-cholinesterases to an intensification and great prolongation of the normally small transmitter action of acetylcholine. Incidentally, it may be noted that such an

* However, cf. **Nachmansohn, D.**[48]
† **Eccles, J. C., B. Katz, & S. W. Kuffler.**[27] FIGURE 5

acetylcholine effect would account well for the latter part of the relatively long transmitter action observed at normal neuro-muscular junctions (5 msec., frog;[52] 6 msec., cat[29].) A similar investigation of the synaptic potentials of motoneurones fails to reveal any such effect of anti-cholinesterases. Hence, it would appear that acetylcholine plays a negligible role at such synapses. These experiments indicate that the synapses of nerve-muscle, sympathetic ganglia, and the spinal cord form a series of decreasing significance for acetylcholine transmission. There has been no reference to other effects of anti-cholinesterases on synaptic transmission, *e.g.*, after-discharge,[6, 7, 23, 27] local contracture,[5, 33, 35] lengthening of refractory period,[29, 51] or catelectrotonic block,[5, 27, 34, 35] because these are all secondarily produced by the prolonged and intensified synaptic potential.

C. Repetitive Synaptic Transmission

There is but little diminution of the synaptic potential set up in curarized mammalian muscle[26] or anesthetized motoneurones[25] by a second pre-synaptic volley, at the shortest intervals after the first. With curarized ganglia, the second potential is usually a little increased.[22] The great increase with the frog's end-plate potential[26, 34, 35, 77] raises a further, as yet insoluble, problem. Since local responses are followed by refractory periods,[42, 70] assumption C of the hypothesis (PART 4) would predict a considerable diminution of a second synaptic potential, at short intervals. However, it must be remembered that the above observations relate to synaptic potentials diminished sufficiently for synaptic blockage, and that small local responses set up much less refractoriness than large responses.[42, 70] A large diminution of synaptic potential is observed, when synaptic transmission is not blocked,[28, 29, 51] and has, hitherto, been attributed to the refractoriness of the post-synaptic cell. Nevertheless, it may be, in part, due to refractoriness, following the postulated large local responses of the specialized post-synaptic membrane (PART 4, C). This latter explanation is supported by the observation that the diminution is lessened by subparalytic curarization.*

8. TESTING THE HYPOTHESIS

It has been shown that the hypothesis gives a satisfactory explanation of all experiments investigating the temporal factors in synaptic transmission. It is otherwise with the intensity factors. Before the hypothesis can be regarded as well established, it has to be shown that

* Eccles, J. C., & S. W. Kuffler[29]: 505.

the currents generated by the pre-synaptic impulses (FIGURE 5) excite the post-synaptic membrane sufficiently to produce the observed synaptic potentials and initiation of impulses. Such an intensity of action is not normally attained, even by the most efficient ephapse,[1, 48] but there are possibly three factors increasing the efficiency of the synapse:

(i) The contact of the pre- and post-synaptic membranes is so intimate[10] that virtually all the current penetrating the former must penetrate the latter (in contrast to the estimated value of one third for the ephapse[48]).

(ii) In the synapse, a large expansion in the area of contact is produced by the swelling and branching of the pre-synaptic terminals; also, with nerve cells, by the multiplicity of closely adjacent synapses.[60]

(iii) If the post-synaptic membrane has special excitatory properties (PART 4, C), the efficiency of electrical excitation may be thereby increased.

By stimulating just beyond the region where a nerve volley is blocked, Hodgkin[41] (cf. also Lorente de Nó[60]) showed that, with critical conditions for blockage, the threshold may be lowered to only 10% of normal, *i.e.*, the nerve volley still provides as much as 90% of the threshold electrical stimulus, beyond the blockage. Hence, there is a high probability that, in the absence of block, it provides an electrical stimulus adequate to excite, *i.e.*, that the transmission of nerve impulses is electrical. It should be possible to apply a similar test to the curarized end-plate of the isolated nerve-muscle fiber preparation. It has, of course, been shown that there is a lowering of threshold, during the end-plate potential,[45, 77] as would be expected for a catelectrotonus, however produced. The present test would explore, instead, the brief interval of pre-synaptic current flow, particularly that preceding the origin of the synaptic potential (*i.e.*, AX in FIGURE 7C). Such a test has already given suggestive results with motoneurones,[58] but, in order to be convincing, there should be an accurate location of the stimulating electrode on the synaptic region of the post-synaptic membrane, and, at present, this seems possible only with the isolated nerve-muscle fiber (cf. Kuffler[54]). An attempt on the whole sartorius was unsuccessful.*

Crucial testing of the hypothesis will also be provided by further pharmacological experiments on synaptic transmission: particularly the action of anti-cholinesterases, and the effects of various background concentrations of acetylcholine, potassium, and calcium. Predictions leading to tests have already been mentioned in the preceding sections.

* Katz, B.[45]: 295.

In addition, mathematical treatment of the flow of penetrating current at a schematized synapse should be possible, and would give more precise predictions for experimental testing. Since the hypothesis is based on the investigations on nerve and muscle fibers outlined in PART 3 (particularly the ephaptic experiments), further developments of this work are of immediate relevance as tests of the hypothesis, providing data on which will depend its development, or modification, or rejection in whole or in part.

Finally, it may be stated that a recommendation for the hypothesis is its systematization of synapses and neuro-muscular junctions in a series (neuro-muscular junctions, ganglionic synapses, and central synapses), exhibiting a progressive replacement of acetylcholine transmission by electrical transmission. To the beginning of such a series could be added those special modifications of ganglionic and neuromuscular synapses, seen, respectively, with the chromaffin organs (suprarenal medulla) and electric organs, where synaptic transmission seems to be wholly due to acetylcholine.

9. SUMMARY

Transmission of impulses across synapses of the spinal cord, sympathetic ganglia, and skeletal muscle, involves a dual problem: (a) the setting up of synaptic, catelectrotonic potentials in the post-synaptic cell; and (b) the initiation of impulses in the post-synaptic cell by such potentials. Evidence is given that, in their present form, both the chemical (acetylcholine) and electrical hypotheses relating to Problem (a) are unsatisfactory. Furthermore, recent experiments are cited which indicate that acetylcholine plays a negligible part as a synaptic transmitter with motoneurones; a subsidiary role with sympathetic ganglia, and possibly also with skeletal muscle. Hence, it is desirable to attempt a more precise formulation of the electrical hypothesis.

The present attempt is based, mainly, on four lines of recent investigation:

(1) *The electrical properties of surface membranes: resistance, electromotive force, capacity, and rectification.* Inductance is neglected, because it is doubtful if the high values of cephalopod axons obtain for vertebrate nerve and muscle.

(2) *The existence of active local responses.* These may be very large in refractory, or anesthetized, or deteriorated nerve.

(3) *The electrical actions occurring across artificial synapses (ephapses).* There are shown to be three main types of ephapses, the synapse

being a special example of the type with an initial anodal and terminal cathodal action; the only type in which a significant excitatory action is exerted.

(4) *Electrical recording* from the isolated neuro-muscular junction shows that the motor end-plate is specialized to give local responses without the all-or-nothing breakdown of propagated impulses. Impulse initiation appears to be produced by a secondary catelectrotonus in the surrounding membrane.

The present hypothesis makes three main assumptions:

A. A schematized formulation is made of the essential geometrical relationship of the membranes of the pre- and post-synaptic elements, as revealed by histological and electrical investigation.

B. The electrical properties of the pre- and post-synaptic surface membranes resemble those observed for peripheral nerve and muscle (see (1), above).

C. The membrane of the immediate post-synaptic region is specialized, so that large and graduated local responses are set up by catelectrotonic polarization (see (2) and (4), above).

On these basic assumptions, it is shown that a pre-synaptic impulse sets up electric currents exerting an initial anodal and later cathodal action on the post-synaptic membrane. The latter action, intensified by rectification, sets up a local response (PART 4, C), which, in turn, acts as a relatively prolonged cathodal focus, from which spreads, electrotonically, the synaptic potential of the effector cell. Finally, the initiation of impulses by this synaptic potential appears to be explicable, simply, as the action of a catelectrotonus.

This hypothesis is shown to offer satisfactory explanations of many fundamental observations on synaptic transmission: irreversibility; synaptic delay; time-course of junctional potential; brief impedance loss at end-plates; dorsal root potentials of the spinal cord; some of which were hitherto inexplicable in detail.

On the other hand, the hypothesis encounters difficulties in explaining the actions of curare and of anti-cholinesterases on synaptic transmission in ganglia and skeletal muscle. The action of curare may be explained, if it is assumed that it depresses the electrical excitability of the post-synaptic membrane, as well as its pharmacological excitability. It is argued that this assumption has not yet been tested. The action of anti-cholinesterases is attributed to the intensification and prolongation of the action of acetylcholine, to which the hypothesis ascribes a subsidiary role, as a transmitter at synapses of ganglia and skeletal muscle. A further difficulty appears to arise in the explana-

tion of rapid, repetitive, synaptic transmission. The postulated local responses should be followed by refractory periods, but a possible explanation is suggested.

The testing of the hypothesis is shown, especially, to concern the further investigation of the special electrical properties assumed for the post-synaptic membrane; also, the attempt to discover how far the postulated electrical actions can account quantitatively for the post-synaptic stimulation. It has been shown (above) that the observed temporal course of the post-synaptic stimulation is satisfactorily explained. In addition, further pharmacological investigation is necessary to test the explanation attributing a subsidiary role to acetylcholine transmission. It is evident that further work on the electrical properties of membranes on local responses, and on ephaptic transmission, will provide additional tests of the hypothesis.

BIBLIOGRAPHY

1. **Arvanitaki, A.**
1942. J. Neurophysiol. **5**: 89.
2. **Barron, D. H., & B. H. C. Matthews**
1938. J. Physiol. **92**: 276.
3. **Bishop, G. H.**
1941. Ann. Rev. Physiol. **3**: 1.
3a. **Bremer, F., V. Bonnet, & J. Moldaver**
1942. J. Arch. Int. Physiol. **52**: 1.
4. **Bronk, D. W.**
1939. J. Neurophysiol. **2**: 380.
5. **Brown, G. L.**
1937. J. Physiol. **89**: 438.
6. **Brown, G. L.**
1937. Physiol. Rev. **17**: 485.
7. **Brown, G. L., H. H. Dale, & W. Feldberg**
1936. J. Physiol. **87**: 394.
8. **Buchthal, F., & J. Lindhard**
1939. The Physiology of Striated Muscle Fiber. Munksgaard. Copenhagen.
9. **Buchthal, F., & J. Lindhard**
1942. Act. Physiol. Skand. **4**: 136.
10. **Cajal, S. R.**
1934. Trav. Lab. Rech. Biol. Madrid **29**: 1.
11. **Clark, A. J.**
1936. Proc. Roy. Soc. London B **121**: 580.
12. **Cole, K. S.**
1941. J. Gen. Physiol. **25**: 29.
13. **Cole, K. S., & R. F. Baker**
1941. J. Gen. Physiol. **24**: 535.
14. **Cole, K. S., & R. F. Baker**
1941. J. Gen. Physiol. **24**: 771.
15. **Cole, K. S., & H. J. Curtis**
1938. J. Gen. Physiol. **22**: 37.
16. **Cole, K. S., & H. J. Curtis**
1939. J. Gen. Physiol. **22**: 649.

17. **Cole, K. S., & H. J. Curtis**
1941. J. Gen. Physiol. **24**: 551.
18. **Cowan, S. L.**
1936. J. Physiol. **88**: 30.
19. **Dale, H. H.**
1937. Harvey Lectures **32**: 229.
20. **Eccles, J. C.**
1936. Ergebn. Physiol. **38**: 339.
21. **Eccles, J. C.**
1939. Ann. Rev. Physiol. **1**: 363.
22. **Eccles, J. C.**
1943. J. Physiol. **101**: 465.
23. **Eccles, J. C.**
1944. J. Physiol. **103**: 27.
24. **Eccles, J. C.**
1945. Nature **156**: 680.
25. **Eccles, J. C.**
1946. J. Neurophysiol. **9**: 87
26. **Eccles, J. C., B. Katz, & S. W. Kuffler**
1941. J. Neurophysiol. **4**: 362.
27. **Eccles, J. C., B. Katz, & S. W. Kuffler**
1942. J. Neurophysiol. **5**: 221, 225, 227–228.
28. **Eccles, J. C., & S. W. Kuffler**
1941. J. Neurophysiol. **4**: 402.
29. **Eccles, J. C., & S. W. Kuffler**
1941. J. Neurophysiol. **4**: 486.
30. **Eccles, J. C., & J. L. Malcolm**
Sent in to J. Neurophysiol.
31. **Eccles, J. C., & J. L. Malcolm**
Unpublished Observations.
32. **Erlanger, J.**
1939. J. Neurophysiol. **2**: 370.
33. **Feng, T. P.**
1937. Chin. J. Physiol. **12**: 331.
34. **Feng, T. P.**
1940. Chin. J. Physiol. **15**: 367.
35. **Feng, T. P.**
1941. Biol. Symp. **3**: 121.
36. **Fillenz, M., & M. Hanafin**
Unpublished Observations.
37. **Gerard, R. W.**
1942. Ann. Rev. Physiol. **4**: 329.
38. **Granit, R., & C. R. Skoglund**
1945. J. Physiol. **103**: 435.
39. **Grundfest, H.**
1932. J. Physiol. **76**: 95.
40. **Grundfest, H.**
1939. Am. J. Physiol. **127**: 252.
41. **Hodgkin, A. L.**
1937. J. Physiol. **90**: 183.
42. **Hodgkin, A. L.**
1938. Proc. Roy. Soc. London B **126**: 87.
43. **Jasper, H. H., & A. M. Monnier**
1938. J. Cell. Comp. Physiol. **11**: 259.
44. **Katz, B.**
1937. Proc. Roy. Soc. London B **124**: 244.

45. **Katz, B.**
1939. J. Physiol. **95**: 286.
46. **Katz, B.**
1939. Electric Excitation of Nerve. Oxford Univ. Press.
47. **Katz, B.**
1942. J. Neurophysiol. **5**: 169.
48. **Katz, B., & O. H. Schmitt**
1939. J. Physiol. **97**: 471.
49. **Katz, B., & O. H. Schmitt**
1942. J. Physiol. **100**: 369.
50. **Kuffler, S. W.**
1942. J. Neurophysiol. **5**: 18.
51. **Kuffler, S. W.**
1942. J. Neurophysiol. **5**: 199.
52. **Kuffler, S. W.**
1942. J. Neurophysiol. **5**: 309.
53. **Kuffler, S. W.**
1943. J. Neurophysiol. **6**: 99.
54. **Kuffler, S. W.**
1945. J. Neurophysiol. **8**: 77.
55. **Langley, J. N.**
1907. J. Physiol. **36**: 347.
56. **Lloyd, D. P. C.**
1943. J. Neurophysiol. **6**: 143.
57. **Lloyd, D. P. C.**
1943. J. Neurophysiol. **6**: 293.
58. **Lorente de Nó, R.**
1935. J. Cell. Comp. Physiol. **7**: 47.
59. **Lorente de Nó, R.**
1938. J. Neurophysiol. **1**: 187.
60. **Lorente de Nó, R.**
1939. J. Neurophysiol. **2**: 402.
61. **Lorente de Nó, R.**
1944. J. Cell. Comp. Physiol. **24**: 85.
62. **MacIntosh, F. C.**
1941. J. Physiol. **99**: 436.
63. **Marmont, G.**
1941. J. Physiol. **133**: 376P.
64. **Marrazzi, A. S., & R. Lorente de Nó**
1944. J. Neurophysiol. **7**: 83
65. **Monnier, A. M.**
1934. L'excitation électrique des tissus. Hermann. Paris
66. **Monnier, A. M.**
1936. Cold Spr. Harb. Symp. **4**: 111.
67. **Nachmansohn, D.**
1940. Yale J. Biol. Med. **12**: 565.
68. **Nachmansohn, D.**
1945. Vitamins and Hormones **3**: 337.
69. **Osterhout, W. J. V., & S. E. Hill**
1930. J. Gen. Physiol. **13**: 547.
70. **Pumphrey, R. J., O. H. Schmitt, & J. Z. Young**
1940. J. Physiol. **98**: 47.
71. **Renshaw, B.**
1940. J. Neurophysiol. **3**: 373.
72. **Renshaw, B.**
1942. J. Neurophysiol. **5**: 235

73. Renshaw, B., & P. O. Therman
1941. Am. J. Physiol. **133**: 93.

74. Rosenblueth, A.
1941 Am. J. Physiol. **132**: 119

75. Rhuston, W. A. H.
1933. J. Physiol. **77**: 337.

76. Rushton, W. A. H.
1937. Proc. Roy. Soc. London B **124**: 201.

77. Schaefer, H., & P. Haass
1939. Pflüg. Arch. ges. Physiol. **242**: 364.

78. Schaefer, H., P. Schölmerich, & P. Haass
1938. Pflüg. Arch. ges. Physiol. **241**: 310.

79. Schmitt, F. O., & O. H. Schmitt
1940. J. Physiol. **98**: 26.

CHEMICAL EXCITATION OF NERVE*

By Frank Brink, Jr.,† Detlev W. Bronk, and Martin G. Larrabee

Eldridge Reeves Johnson Research Foundation, University of Pennsylvania, Philadelphia, Pennsylvania

One of the noteworthy characteristics of neurones is their sensitivity to changes in the chemical environment. Even within the relatively protected interior of the body, the properties of nerves are subject to modification by variations in the composition of the body fluids. Indeed, the alterations of irritability and the trains of nerve impulses, which are the result of changes in the chemical environment, are among the most important factors involved in the regulation of the activity of the organism. This is one of the significant reasons for studying the chemical activation of nerve. A second reason derives from the current interest in the role of chemical agents in the mechanism of synaptic transmission. Furthermore, the investigation of the effects of various chemical agents is one of the most fruitful sources of information regarding the role of the several chemical components of the nerve structure and of the chemical processes involved in nervous action.

This last consideration suggests that the most significant chemical agents for use in the study of the processes of activation are those which have an important part in the normal structure of nerve. Calcium is such an element. Potassium is another; it modifies the action of calcium, to which it is closely related in the regulation of nerve action, and it has a marked influence on the electric potential difference across the interfaces at which the nerve comes in contact with its environment. Finally, the effects of acetylcholine on the initiation and conduction of the nerve impulse make an important and timely subject for investigation in such a study as this. It is with the effects of these agents that we shall be primarily concerned. There are others of significance for a general study of this problem, but from these three we can derive many of the basic phenomena involved in chemical excitation.

I

The changes in the functional characteristics of a nerve caused by an alteration of its chemical environment are due to the consequent

* The experimental work reported here has been generously supported by grants from the Supreme Council, Scottish Rite Masons, and from the American Philosophical Society.

† Fellow of the Lalor Foundation.

changes in the chemical constitution of the cell or axon. Thus, the increased irritability that is induced by surrounding a nerve with a calcium-deficient fluid follows a decrease of calcium within the nerve. In order that the effects of such changes may be investigated, it is convenient to have available a solution which will maintain nerve in a stable functional state for long periods of time, and to which the effects of other environmental solutions may be referred. It is customary to choose for reference a solution having a salt content and pH approximating that of the animal's body fluids. By direct test, it has been found that a frog nerve can be kept in a solution at pH 7.2 (phosphate buffer) containing sodium chloride (116 mM), potassium chloride (2.0 mM), and calcium chloride (1.8 mM) for many hours, with no significant change in excitability or in rate of aerobic oxidation. Squid nerve, which we have also employed, maintains a similar stable functional state in sea water (Woods Hole) at pH 8.0 or in a solution containing sodium chloride (405 mM), potassium chloride (11 mM), and calcium chloride (70 mM). Modifications of these solutions have been used as the experimental means of chemical activation.

The calcium ion concentration of the environmental fluid is especially important in determining the excitability of nerve. This familiar phenomenon (cf., *e.g.*, Misske[1]) can be studied quantitatively and under quickly reversible conditions in squid giant axons or in bundles of frog axons from which the perineurium has been removed. Under those circumstances, diffusion equilibrium between the axons and the surrounding fluid is attained relatively quickly. In FIGURE 1, the threshold strength of direct current necessary to initiate an impulse, which is the rheobase, is plotted as a function of the concentration of calcium chloride in the fluid bathing a giant axon of the squid. A similar relation is obtained for the α fibers in a frog sciatic nerve (FIGURE 2).

The increased excitability produced by the action of solutions having a low concentration of calcium chloride is presumably due to the diffusion of Ca^{++} from the cell structure. Indeed, Tipton[2] has shown by chemical analysis that as much as 40 per cent of the total calcium of frog nerve is in diffusion equilibrium with the surrounding fluid, some of this diffusible calcium being in the cell phase. His evidence for intracellular precipitation of added calcium is a further indication that changes in the calcium chloride content of the bathing fluids lead to changes in the cellular content. The spatial distribution of these changes in cellular calcium are unknown.

When frog nerve is equilibrated with solutions containing from 1.0 mM to 0.3 mM calcium chloride, or when squid nerve is equilibrated

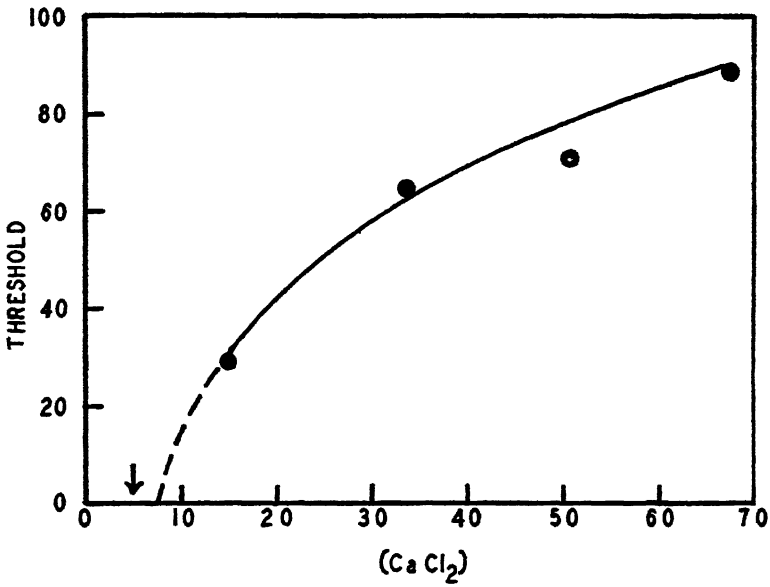


FIGURE 1. Threshold of giant axon of Squid in isotonic solutions containing various concentrations of $CaCl_2$. Threshold (rheobase) as per cent of threshold of nerve in Woods Hole sea water. Concentration in millimoles per liter. The arrow indicates concentration at which spontaneous activity began.

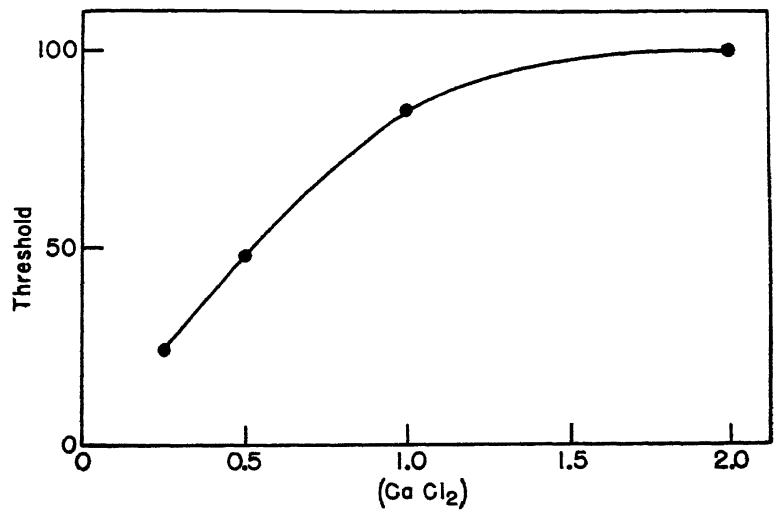


FIGURE 2. Threshold of α fibers of sciatic nerve of frog with various concentrations (millimoles per liter) of $CaCl_2$ in bathing fluid. Threshold as per cent of rheobase of nerve in reference solution described in text.

with solutions containing from about 70 mM to 10 mM calcium chloride, there is a certain degree of irritability corresponding to each concentration. This is measured as the minimal strength of current necessary to initiate a conducted impulse. It may be thought of as an index of the stability of the excitable portion of the nerve structure. We shall subsequently refer to the fact that the rate of oxygen consumption of nerve is also modified by changes in its calcium content. Here, it is pertinent to remark that the variations of oxygen consumption occur within this same range of calcium concentrations in which there are measurable changes of irritability. In the case of frog nerve, moderate increases in calcium above 2.0 mM do not cause a further appreciable decrease of irritability nor a further decrease of oxygen consumption. At very much higher concentrations, above 15 mM, the irritability again decreases,[3] and there is a further fall in the oxygen consumption.[4] If the concentration of calcium be lowered beyond 0.3 mM, or 10 mM in the case of squid nerve, it is no longer possible to measure the irritability in terms of the strength of current necessary to initiate an impulse. At these levels, the nerve structure has been so much modified, its stability lowered so much, that it goes through periodically-recurring cycles of change, with consequent, self-initiated trains of propagated impulses.[5]

II

The response of nerve to the exciting action of an electric current can be studied in a nerve trunk or in a bundle of fibers. The stimulus is under the control of the experimenter, and all of the fibers are excited simultaneously. Accordingly, the action potential recorded from the aggregate of fibers of a given type is a fairly accurate representation of the sequence of events in each fiber, provided temporal dispersion, due to differences in conduction velocity, is avoided.

The situation is quite different in the case of chemical excitation. The altered chemical environment modifies the properties of the fibers, so that a sequence of cyclic events develops in each fiber, with a frequency that is determined by the characteristics of the fiber. Because these intrinsic characteristics differ, the frequency of the impulses discharged from a chemically treated region varies from fiber to fiber. Furthermore, the properties of the fiber may change from moment to moment, so that the sequence of impulses is not truly periodic. Finally, the times of initiation of impulses in one fiber are independent of the timing of these events in the other fibers, in contrast to the externally determined synchronization imposed by electric stimuli. Because of

these considerations, the action potentials recorded from a bundle of numerous fibers reveal little of what is occurring in the individual units (FIGURE 3). Under these conditions, the investigation of the processes of excitation and response encounters the same difficulties experienced in the study of groups of sensory endings or motor nerve cells (Adrian;[6] Adrian and Zotterman;[7] Adrian and Bronk[8]).

The solution here is the same as there: that is, to isolate and measure the activity in a single fiber. Only when this is done can one observe the more or less rhythmic train of impulses discharged from the chemically modified region (FIGURE 3). The difficulties inherent in this

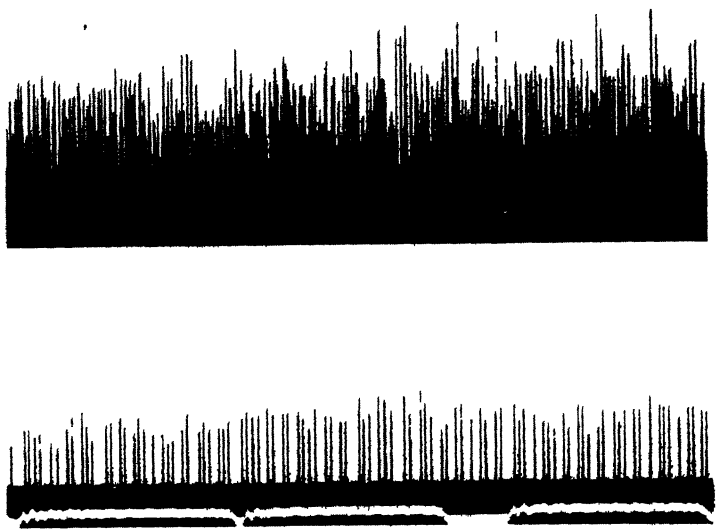


FIGURE 3. Above Impulses recorded from branch of sciatic nerve of frog stimulated by topical application of isotonic sodium citrate. Below Impulses recorded from single α fiber dissected from this nerve Time in 1/5 seconds.

experimental procedure partly explain the relative paucity of our knowledge regarding the nature of chemical excitation.

It is worthy of emphasis that an axon possesses the capacity (as does a cell body or sensory ending) for transforming the continuous environmental action of a physical or chemical agent into a series of recurring events which are made manifest as nerve impulses.

It has been said that the frequency of impulses developed varies from fiber to fiber, and depends upon the intrinsic characteristics of each

fiber. Obviously, these properties will be modified by the chemical excitant, qualitatively by the nature of the chemical change, and to a degree that is determined by the amount of substance added to, or removed from, the nerve structure. Therefore, it is not surprising to find that the number of impulses discharged per second from the treated region depends upon the concentration of the calcium ion, as shown in FIGURE 4.

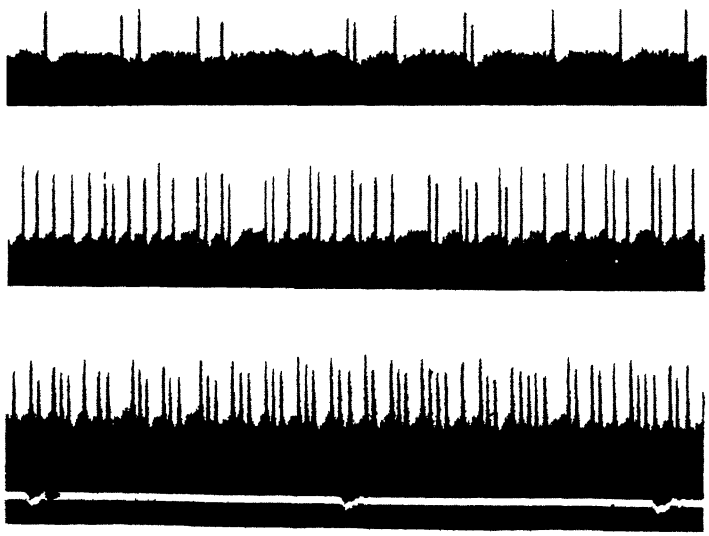


FIGURE 4. The average frequency of impulses recorded from a single fiber depends upon the concentration of calcium ions. Upper record, concentration of Ca ++ is 0.3 mM, middle, 0.1 mM; lower, no calcium. In this experiment, the 0.1 mM calcium was applied first, then the 0.3 mM, and finally the calcium-free solution. Time in 1/5 seconds.

The frequency of impulses initiated by a given reduction of calcium ion concentration, or by other chemically stimulating media, also depends upon the previous duration of the chemical action. The electric threshold begins to fall almost at once after the application of the solution (FIGURE 5). Further time is required for changes in the intracellular processes which must precede the development of conducted impulses. Indeed, the first impulse may not develop for some minutes, and then at a time when the threshold has fallen to zero. This same gradual loss of stability, continuing further, is manifest in the progressive increase in the average frequency of impulses.

When calcium is removed from a nerve, by diffusion into a solution containing less than the normal amount of calcium chloride, the dis-

charge of impulses referred to above begins slowly. If the calcium ion concentration is reduced, by adding a calcium-binding agent such as sodium citrate, the discharge begins more quickly and continues longer. Because of this, we have used for many of our experiments a stimulating fluid which contained sodium citrate in place of some of the sodium chloride. When the calcium ion concentration is thus reduced below about 0.4 mM, impulses are initiated. This is the concen-

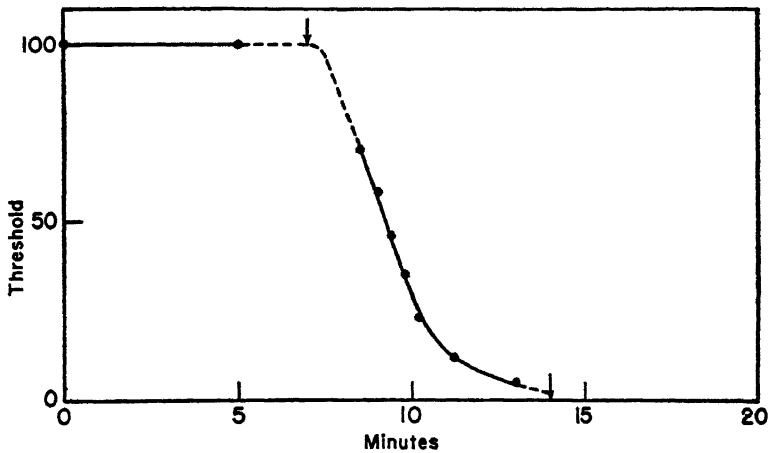


FIGURE 5. The threshold of an axon decreases with time after topical application (upper arrow) of a solution which lowers the calcium ion concentration. Repetitive activity begins at 14 min. after rheobase has decreased below five per cent of its initial value. Threshold measured with cathode on calcium-deficient region of nerve.

tration level for activation, whether the calcium be removed by the action of citrate or by the simple process of diffusion. This, and other evidence, suggest that a principal factor in citrate excitation is the lowered calcium ion concentration.

We have already stated that the frequency of impulses initiated by the removal of calcium is, at any time, dependent upon the duration of the previous action of the stimulating fluid. The time-course of development of the impulse discharge is also largely influenced by the previous chemical treatment of the nerve. Usually, the impulses begin to occur, at random intervals, when the rheobase has fallen to about 5 per cent of its initial value. Thereafter, the impulses are discharged in groups, which gradually merge into a more or less regular train when the frequency reaches about 150 per second. Such a gradual increase in the frequency of impulses is shown in one of the curves of FIGURE 6.

The development of activity is not always so gradual. Sometimes, the initial frequency may be high, and then decline to a lower level that is sustained for some hours, with the development of hundreds of

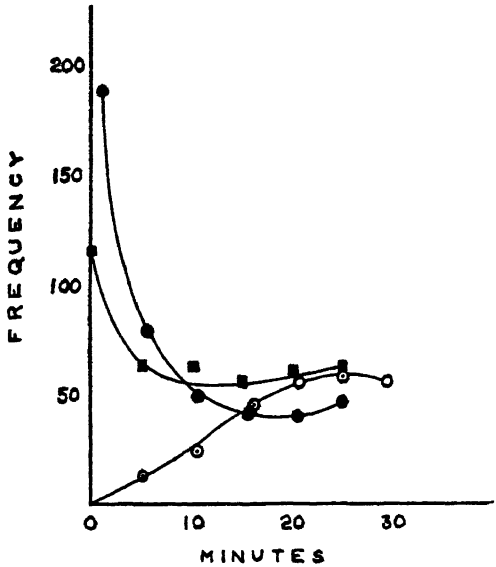


FIGURE 6 Time-course of discharge (impulses per second in a single fiber), after topical application of isotonic solution containing sodium citrate (35 mM) and sodium chloride
Open circles gradual increase of frequency during first period of chemical excitation
Squares response to second application of same solution, after intervening 2 hours in reference fluid
Filled circles response to a third application, after another hour in Ringer's fluid.

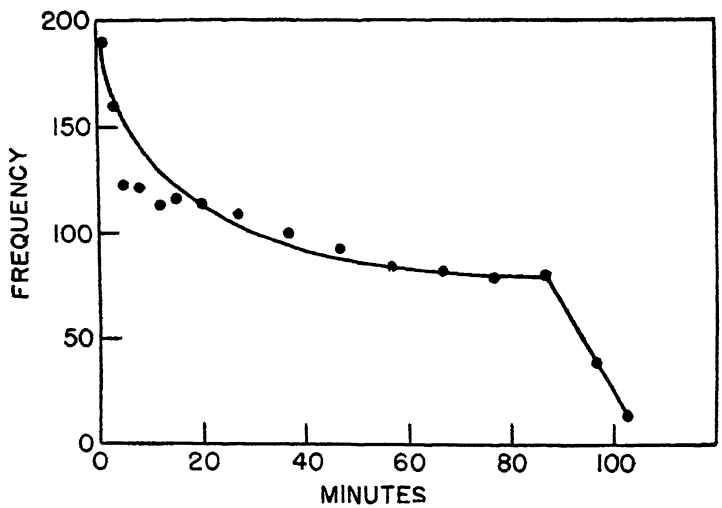


FIGURE 7. The discharge of impulses continuing at high frequency for many minutes during topical application of isotonic sodium citrate. Approximately 540,000 impulses were produced by the single fiber, during the activity plotted in this figure.

thousands of impulses (FIGURE 7). This initial high frequency discharge usually occurs after a previous period of chemical excitation that had been arrested by the restoration of calcium (FIGURE 6). It also occurs during the actual restoration of calcium.[9]

A nerve which has been once modified by the withdrawal of calcium continues to give such a response of high initial frequency to a successive activation, even though it has been in unmodified Ringer's fluid for many hours (FIGURE 6). Whether the impulses start at a high frequency that declines, or whether the frequency gradually increases, the final, sustained average frequency is about the same for a given stimulating fluid. This frequency is, to an important degree, determined by the calcium content of the nerve, and our experiments also suggest that it is, in part, dependent upon the rate of removal of calcium. This latter factor may be especially important in the determination of the transient changes of frequency.

III

The resting metabolism of nerve has long been thought of as necessary for maintaining the organization of its unstable structure against the tendency of the structure to become disorganized. In accordance with this view, the less stable structure resulting from the withdrawal of calcium should have a higher metabolic requirement for its maintenance. This increased metabolism has been observed.[10] We, too, have made such measurements of the oxygen consumption of nerve from which varying amounts of calcium have been withdrawn, while, at the same time, measuring the excitability and recording any impulses that were initiated.[11] Alterations of calcium content in the nerve, sufficient to cause a lowered threshold to electric stimuli, but insufficient to cause the rhythmic discharge of impulses, induce an increased oxygen consumption (FIGURES 8 and 9). As the calcium content is further reduced, the oxygen consumption increases still more. Finally, a level of calcium content may be reached which is sufficiently low to cause the rhythmic discharge of impulses, and associated with this calcium content there is a still higher oxygen consumption (FIGURE 8).

This progressive increase of oxygen consumption of nerve with decreasing concentrations of calcium, starting at calcium levels too great to permit the development of spontaneous activity, raises the question as to the meaning of the term, "resting oxygen consumption." The oxygen consumption of axons which are resting, in the sense of not conducting impulses, may be quite different in different chemical environments.

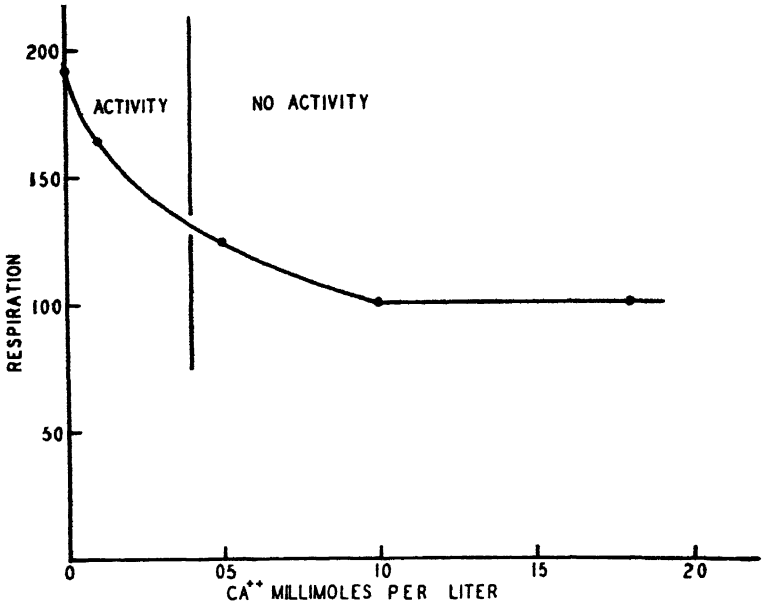


FIGURE 8. The respiration (as per cent of value in reference solution) of a frog sciatic nerve increasing when the concentration of calcium chloride in bathing fluid is below 1.0 mM. There is an appreciable increase at a concentration which is not low enough to initiate impulses.

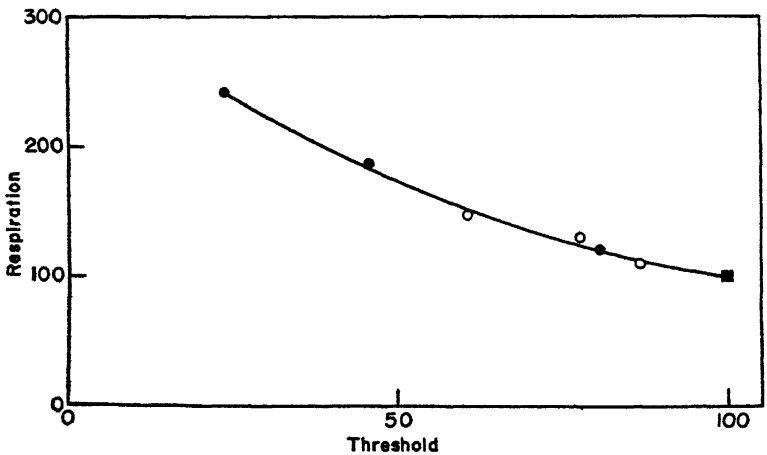


FIGURE 9. Respiration of frog sciatic nerve in relation to threshold of α fibers in same nerve. Changes produced by equilibration in isosmotic solutions containing 1.0, 0.5, and 0.25 mM $CaCl_2$. Both respiration and threshold (rheobase) are expressed as per cent of their values (marked by solid square) in reference solution (containing 2.0 mM $CaCl_2$).

Open circles: A nerve showing a small per cent change in respiration has a correspondingly small per cent change in rheobase.

Solid circles: Another nerve, showing a large per cent change in respiration, has a large per cent change in rheobase.

An increased rate of oxidation is essential for the initiation of impulses. That is shown by the effects of an oxidation-inhibiting agent, such as sodium azide. A portion of a nerve trunk was placed in sodium citrate, and impulses which were thus developed were recorded in one of the fibers coming from the chemically activated region. Sodium azide was then applied to the citrate-treated portion of the nerve, in a concentration that was sufficient to suppress the chemical excitation. For this, a concentration of azide which restored the oxidation to a normal rate was adequate (FIGURE 10). After the rhythmic discharge

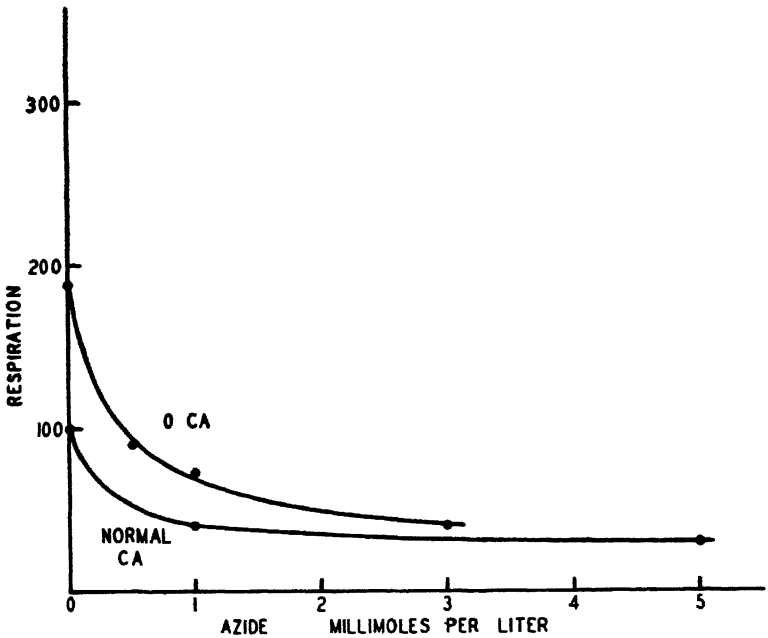


FIGURE 10. The respiration of a calcium-deficient nerve (per cent of value in reference solution) suppressed by sodium azide, as is the respiration of a normal nerve.

of impulses had thus been inhibited, it was still possible, for several hours to send a high frequency train of impulses, initiated by electric stimuli, through the calcium-deficient and azide-treated length of nerve.

The initiation of impulses by a calcium-deficient region of nerve requires a higher rate of oxidation than is necessary for the maintenance of the capacity of such a calcium-deficient region to conduct impulses. Certain specific chemical changes in the constitution of nerve and increased metabolic rate both appear to be necessary for the initiation of rhythmic activity. One without the other is an inadequate condition for self-excitation.

Although there are many instances of a close parallelism between the stability of nerve, as measured by its electric threshold or by the spontaneous development of impulses, and its rate of respiration, there are exceptions. For instance, the rhythmic activity may be abolished by the application of potassium chloride, which, at the same time, increases the rate of respiration. Another instance of such a lack of parallelism is revealed when a nerve is returned to its normal fluid environment, after treatment with sodium citrate. The rhythmic activity is promptly suppressed, and the threshold becomes normal, long before there is a corresponding recovery of the original, normal rate of respiration.

IV

There has been a persistent notion that the initiation of trains of impulses from a chemically activated portion of an axon or from a sense organ under a constant stimulus is due to a gradient of electric potential at the site where the impulses originate. Indeed, Adrian[13] found that there was a gradient of 10 mV between a normal portion of nerve and an injured region from which impulses were discharged. Accordingly, he attributed the excitation to this demarcation potential. Furthermore, Erlanger and Blair[14] and Fessard[15] caused the rhythmic discharge of impulses by the passage of constant currents. Finally, Katz[16] and Arvanitaki[17] found that the duration of such an electrically induced repetitive discharge could be much prolonged by reducing the calcium content of the nerve.

Because of these considerations, we have carefully searched for some causal relation between a potential gradient developed at the site of calcium removal and the chemical initiation of impulses. To do this, one of a pair of non-polarizable electrodes was placed in contact with the chemically altered region of the nerve; the second was in contact with an adjacent, untreated portion. We have found that the activity develops in the calcium-deficient part of the nerve, without the appearance of an appreciable longitudinal potential gradient.[18] However, this part of the nerve is very sensitive to weak currents. It is, therefore, possible that potential differences of less than a millivolt might be involved in the mechanism of chemical excitation.

Accordingly, experiments were devised to study quantitatively the relation between the frequency of conducted impulses and changes in membrane polarization produced in the hyper-excitable portion of the axon by certain additional chemical agents. Increasing the proportion of potassium chloride in the solution of sodium citrate used to excite

the nerve makes the calcium-deficient region negative to the adjacent parts of the cell. Under these conditions, the associated current flow is inward across the plasma membrane in the hyper-excitable region of the axon. The conducted impulses still occur, but at a reduced frequency, compared with the activity initiated by a solution containing less potassium chloride. The frequency of response is lower, the

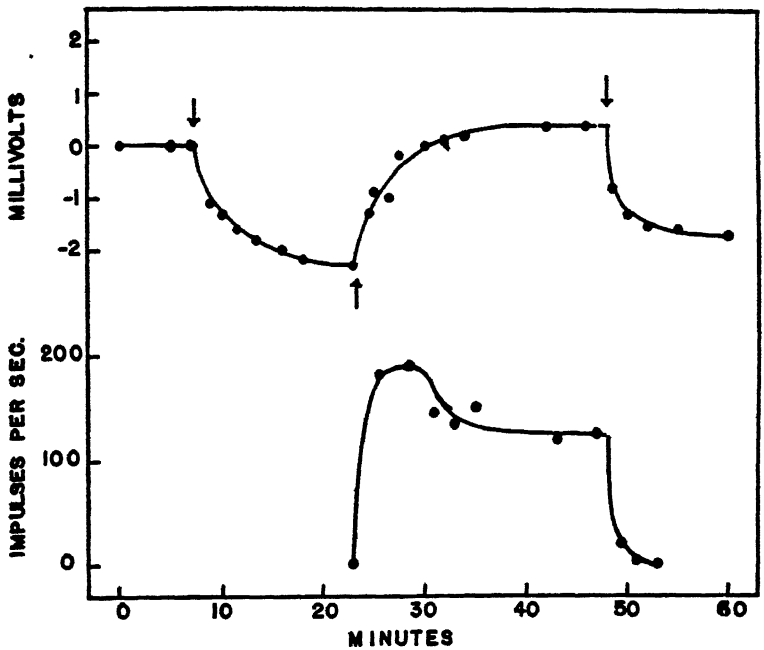


FIGURE 11. The depression by KCl of the activity produced by topical application of isotonic sodium citrate is related to the associated decrease in membrane polarization.

Upper curve is potential difference between treated region of nerve and normal part of nerve. At the first arrow, the solution was changed from isotonic sodium citrate to one containing the same amount of citrate, but with 15 mM K in place of some of the sodium. At the second arrow, the nerve was returned to a solution containing isotonic sodium citrate without potassium. At the third arrow, the solution containing potassium was again applied to the nerve.

The lower curve shows change in frequency of response recorded from a single fiber in this nerve. A similar suppression of activity during the first cycle of depolarization was observed, but not recorded.

higher the concentration of potassium chloride, and, therefore, decreases, as the degree of depolarization increases. When the calcium-deficient part of the cell is made sufficiently negative, the activity is suppressed, but it begins again as this depolarization is removed by washing out the potassium chloride (FIGURE 11). Conversely, if this region of the nerve is made positive to adjacent parts, as by a solution of sodium thiocyanate, the frequency of the impulses is increased.

Thus, small differences of potential between a normal and hyper-irritable region of an axon modify the frequency of the impulses dis-

charged from that region, despite the fact that the impulses originate without the mediation of such potential gradients. The influence of these gradients on the excitability of calcium-deficient nerve, as measured by changes in the frequency of impulse discharge, is in agreement with the usual effects of current flow from an external source: depression at the anode and excitation at the cathode.

Although the chemical excitation of nerve does not depend upon the development of a steady current flow, the discharge of chemically initiated impulses can be modified by an externally imposed potential

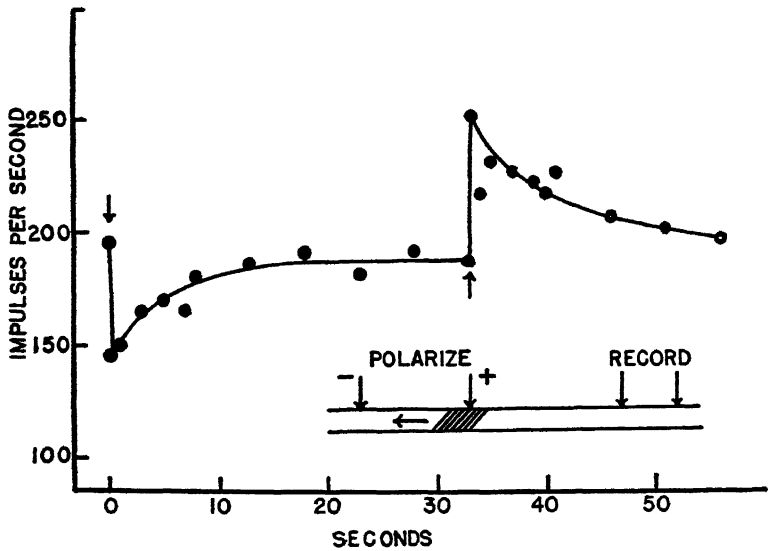


FIGURE 12 The frequency of response in a single fiber stimulated by topical application of isotonic sodium citrate is reduced when the treated region is anodally polarized (first arrow). When circuit is opened (second arrow), there is a transient increase in response. The current passed into nerve in the treated region as shown in the diagram

gradient. This was first reported by Fessard,[19] who observed such an effect in crab nerves which had been excited to activity by the application of alcohol or sodium thiocyanate. To study this problem further, we have passed a polarizing current through a calcium-depleted region of frog nerve, during the period of constant frequency of discharge. When the direction of current flow is such that it enters the nerve fiber in the chemically activated region, the average frequency of impulses is reduced for a brief time. As shown in FIGURE 12, only a slight depression continues after a few seconds. When the polarizing current is terminated, there is a temporary increase in the frequency of impulses from the chemically activated region, followed by a return to the frequency that preceded the beginning of the current flow.

If the current is in the opposite direction, so that it flows out of the fiber in the calcium-deficient region, the sequence of frequency changes is reversed. Then, as the current starts, there is a transient increase of frequency, followed by a temporary depression when the current is interrupted.

Still obscure are the cellular mechanisms which account for an increase in the frequency of impulses from calcium-deficient nerve, when positively charged ions move outward across the fiber interface, or for

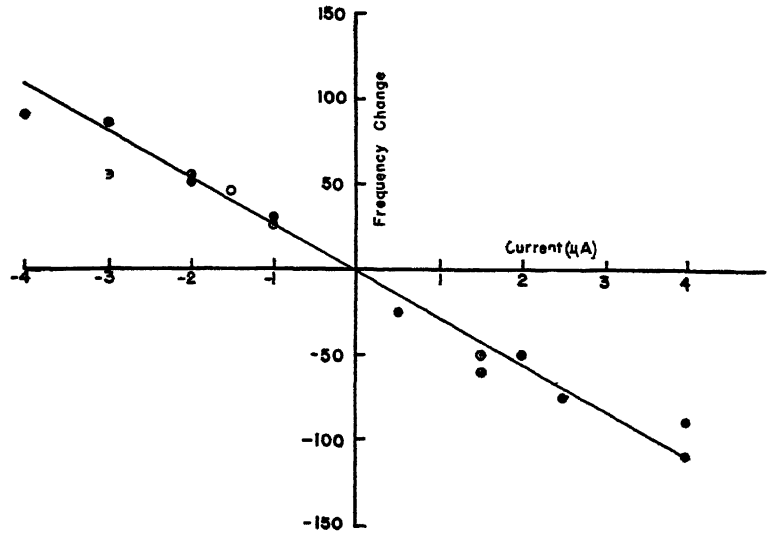


FIGURE 13 The maximum change in frequency (impulses per second) in a chemically excited frog α fiber is proportional to the magnitude of polarizing current (in microamperes). Nerve excited as described in FIGURE 12
Positive current, anode in the treated region; negative current cathode in treated region.

a decrease of frequency, when the ionic movement is reversed. The effects are, however, consistent with the long-established fact that a cathodally polarized region of nerve is more irritable, while anodally polarized nerve is less irritable. Accordingly, the effects are also in agreement with the view that an agent which reduces the stability of the nerve structure increases the frequency of chemically induced activity.

The magnitude of the transient increase or decrease in the frequency of impulses, caused by the passage of an electric current through a chemically activated nerve, depends upon the strength of current. FIGURE 13 shows that there is, indeed, a linear relationship between the current strength and the maximal increase or decrease of impulse frequency caused by the current flow, within certain limits. This figure

also reveals the significant fact that there is, apparently, no minimal current strength that must be exceeded before the rhythmic discharge from the chemically sensitized nerve is modified. Any change in the direct current, no matter how small, flowing across the membrane of these chemically modified nerves, alters the rhythmic activity of the nerve and is reflected in the altered frequency of the propagated impulses. This is in contrast to the limiting threshold of current strength necessary for the excitation of a conducted impulse in a nerve with normal calcium content.

V

We have already said that potassium chloride causes a decrease in the frequency of the impulses developed in a calcium-deficient portion of

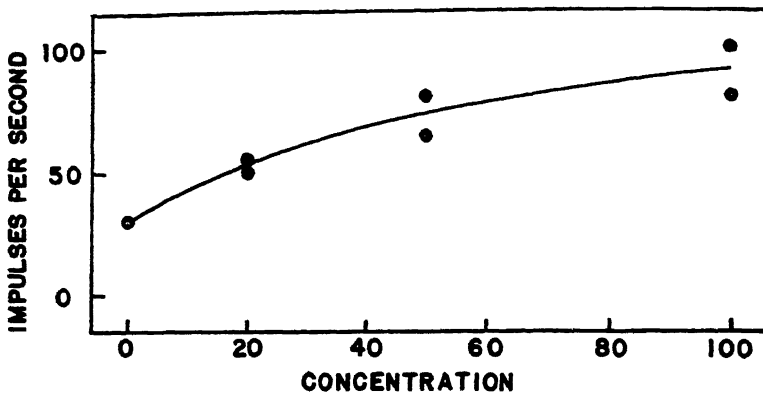


FIGURE 14. Frequency of response due to action of isotonic solutions containing various proportions of sodium chloride and sodium citrate. Concentration in terms of per cent of isotonic solution of sodium citrate.

a nerve. This is the effect of potassium chloride when its action has reached a steady state, but it is preceded by a transient increase in the number of impulses discharged per second. This stimulating action of potassium chloride also occurs in nerves with normal calcium content, but there it is of even shorter duration. Both the maximum frequency and the duration of the impulse discharge, caused by an increase of potassium chloride, are greater, the lower the calcium content of the fluid bathing the nerve.[20] Thus, the removal of calcium from a nerve makes it more sensitive to the transient stimulating action of a moderate increase in the concentration of potassium chloride.

This increased sensitivity of calcium-deficient nerve to other chemical agents is further revealed in the experiments shown in FIGURE 14. There, the nerve was made active by bathing a portion of it in isotonic

sodium chloride. When the frequency of response had become constant, the sodium chloride solution was replaced by one in which some of the sodium chloride had been replaced by sodium citrate. The frequency of impulse discharge then increased, as is shown. With each further increase in the proportion of sodium citrate, there was a further increase in the average number of impulses per second. When the nerve was subsequently returned to solutions containing successively smaller concentrations of citrate, there was a parallel decrease in the impulse frequency. Finally, in isotonic sodium chloride, the initial low degree of activity was resumed. This decrease in average frequency, associated with the return to isotonic sodium chloride, is obviously not due to a restoration of calcium to the nerve. It seems probable, therefore, that a nerve made active by removal of calcium is sensitive to changes in the concentration of the citrate ion. This contrasts with the previously mentioned lack of effect of citrate upon a nerve in the presence of Ringer's proportion of calcium ions.

In a similar manner, the stimulating action of sodium thiocyanate is enhanced by first removing some of the calcium from the nerve. Also, tetraethyl ammonium chloride will stimulate frog nerve,[21] and we have found that its effectiveness in initiating impulses is greater, if the axon is sensitized by preliminary removal of some of the calcium.

Another quaternary ammonium salt of interest in this discussion is acetylcholine. Lorente de Nó has shown[22] that it does not alter the membrane potential of frog nerve, even in massive concentrations. On the other hand, Nachmansohn argues for the possibility of such an action, on the grounds that cholinesterase-inhibiting agents, which should permit the accumulation of acetylcholine, do cause depolarization of squid nerve.[23] In our experience, acetylcholine does not induce a discharge of impulses when it is applied to the axons of a peripheral frog nerve trunk, and we have not been able to increase the frequency of chemically excited impulses by adding acetylcholine to the calcium-deficient fluid. Also, we have investigated the effects of this substance on mammalian nerve, by perfusing the stellate ganglion of a cat. In no case have we found any evidence that impulses are thus initiated in the pre-synaptic fibers within the ganglion, even though as much as 500 micro-grams of acetylcholine were added to each cc. of perfusion fluid. This was determined by observing that no impulses were discharged over the fibers of the preganglionic trunk (FIGURE 15). Finally, there remains the contrasting and significant observation, that much lower concentrations of acetylcholine do cause the discharge of rhythmically recurring impulses in the post-synaptic neurons. The cell bodies or the

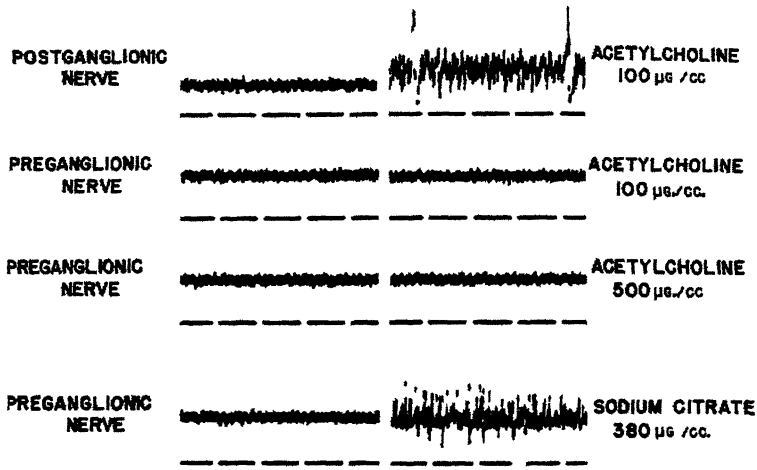


FIGURE 15 Responses recorded from the preganglionic and postganglionic nerves of a cat's stellate sympathetic ganglion, during perfusion with acetylcholine and sodium citrate Control records in the absence of a chemical excitant in the left hand column Time 0 1 sec

immediately contiguous portions of their axons differ from axons in general in some way that makes them sensitive to the action of this agent.

This is a striking example of a specificity of nerve structure involved in the process of chemical excitation. It has been suggested that the basis of this differential action is the presence or absence of a myelin sheath that would prevent the rapid penetration of the acetylcholine.[23] That is not likely to be the explanation of the contrasting effects in the experiments just reported, for the terminal portions of the pre-synaptic fibers within the ganglion are considered to be non-myelinated,[24] as are the post-synaptic neurons.

Certain chemical agents, such as acetylcholine, are highly specific, with regard to the type of nerve structure they excite. Others, of which citrate and calcium-deficient solutions are examples, are quite general in their action.[25] Thus, a reduction of calcium ions in the perfusing fluid, or the addition of sodium citrate, causes the discharge of recurring impulses in both the pre-synaptic and the post-synaptic neurons, as in axons generally (FIGURE 15).

The character of the response of ganglion cells to acetylcholine is, in many respects, analogous to the response of peripheral axons to chemical excitation. For each cell, there is a threshold concentration that must be exceeded before impulses are developed This threshold differs from cell to cell, but it is usually less than 25 micrograms of acetyl-

choline per cc., during continuous perfusion with solutions containing no inhibitor of cholinesterase. When the threshold for a cell is exceeded, it discharges impulses with a regular rhythm, and for extended periods that we have observed to be as long as an hour. The frequency of this

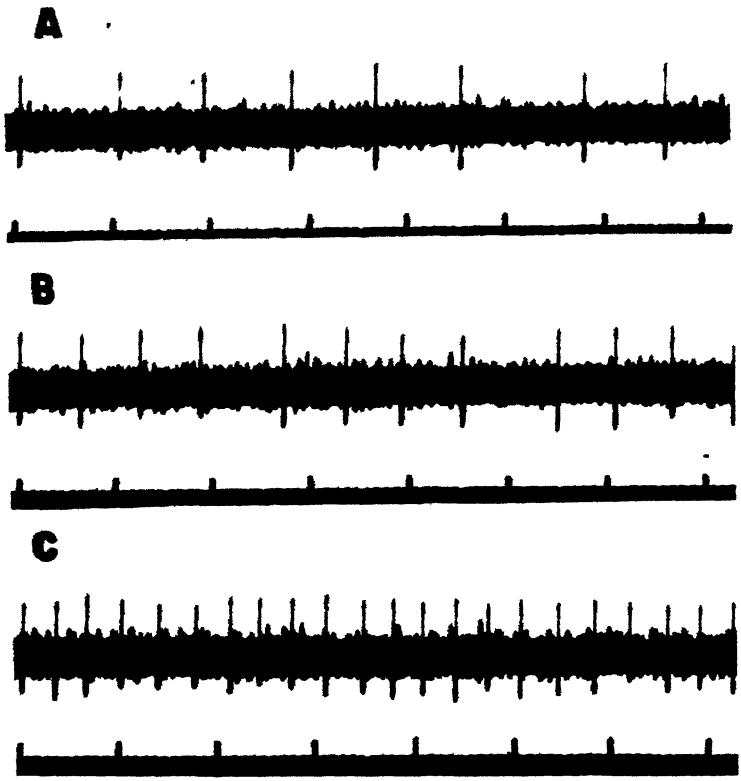


FIGURE 16 Impulses discharged from a single sympathetic ganglion cell in response to acetylcholine in concentrations of A 25; B 50, C 100 micrograms per cc Cat's stellate ganglion perfused with a modified Ringer's fluid containing acetylcholine, but no inhibitor of cholinesterase Impulses recorded from a fine strand of the postganglionic nerve Time in seconds

discharge increases with increased concentrations of acetylcholine (FIGURE 16). Finally, at concentrations of about 200 micrograms per cc., the excitatory action ceases, the discharge of impulses is arrested, and the ganglion cells cannot be stimulated by volleys of preganglionic impulses.

The response of nerve to chemical excitants depends upon the totality of environmental agents. This has been emphasized before. It is a fact that is illustrated by the effects of the combined action of acetyl-

choline and other chemical agents on ganglion cells. For example, the frequency of impulses initiated by a certain concentration of acetylcholine may be reduced by increasing the concentration of calcium in the perfusion fluid, or by reducing the concentration of potassium (FIGURE 17). Conversely, the frequency of discharge may be increased by

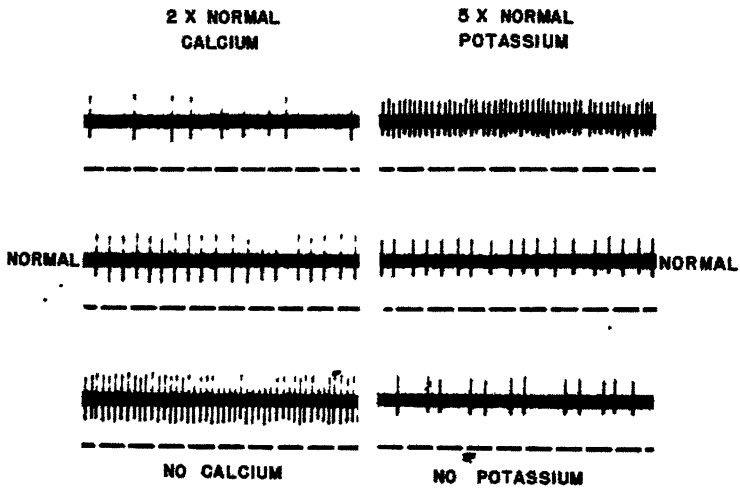


FIGURE 17. Impulses discharged from a single cell in a cat's stellate ganglion, during perfusion with six different solutions, all containing the same amounts of acetylcholine (40 μg 1 cc.), but different concentrations of calcium and potassium. Time in seconds

lowering the concentration of calcium or by augmenting the concentration of potassium. These effects of calcium and potassium on the rhythmic action initiated by another chemical agent might be anticipated from our knowledge of their effects on the electrical excitability of axons. It is, perhaps, worthy of comment that, regardless of whether the acetylcholine, calcium, and potassium act upon the same or different parts of the irritable mechanism, their combined effects become manifest in a modification of the rhythmic process which initiates the propagated impulses.

This repetitive process, which is a latent characteristic of nerve, is revealed in the discharge of impulses initiated by sensory stimulation or by chemical action. It is also to be observed in the periodic activity of nerve cells which are excited by the arrival of impulses in adjacent pre-synaptic endings. When the preganglionic fibers entering a sympathetic ganglion are stimulated by repetitive electrical shocks of high frequency, the cells discharge repetitively, but at a much lower fre-

quency. Furthermore, the cells discharge their impulses in no fixed temporal relation to the incoming, excitatory impulses. In short, the cells which are activated through the pre-synaptic endings initiate impulses at a frequency that depends upon the characteristics of each cell, as well as upon the frequency and the number of pre-synaptic im-

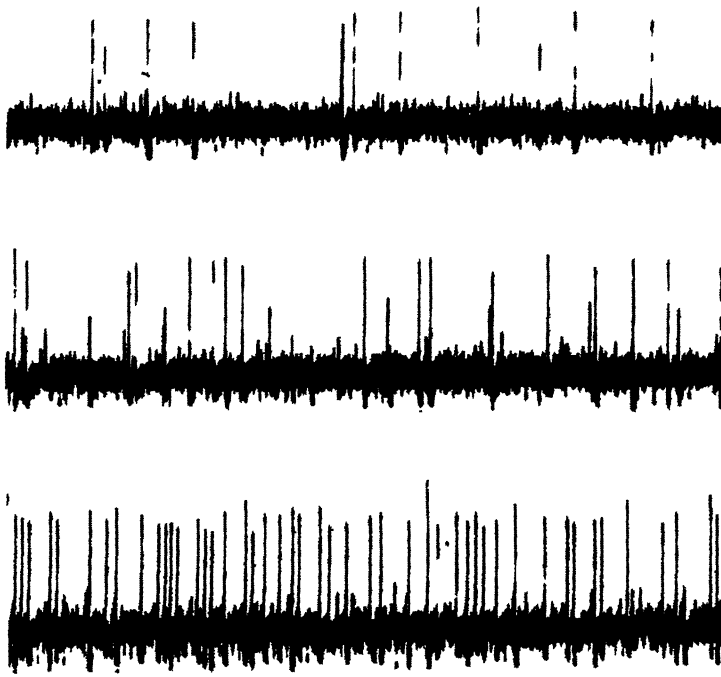


FIGURE 18 Impulses discharged by a few sympathetic ganglion cells in response to stimulation of the preganglionic nerve at a frequency of 50 per second. The ganglion was perfused with a modified Ringer's solution containing various amounts of calcium chloride
Uppermost record, 4 4 mM, middle record, 2 2 mM; bottom record, 1 1 mM The middle record represents the normal level of calcium
Time in 0 1 seconds

pulses.[27] If these cellular characteristics are modified by any means (by nerve impulses or by chemical agents), the rhythmic processes are altered, and this modifies the frequency of their action.

Such a modification of the rhythmic response of a nerve cell to neural activation can be accomplished, as would be expected, by varying the concentrations of calcium ions in the synaptic regions (FIGURE 18). If activity is excited in a ganglion cell by trains of pre-synaptic impulses, the frequency is decreased by raising the concentration of calcium in the perfusion fluid. Alternatively, the frequency of the impulses dis-

charged from the ganglion cells can be increased by decreasing the calcium concentration. It must be said that a further reduction of calcium in the perfusing solution may cause a complete block of ganglion cell excitation by impulses in the pre-synaptic fibers.

VI

Some evidence regarding the nature of the cellular events which cause the more or less rhythmic discharge of impulses from a chemically excited region of nerve can be derived from a consideration of the temporal distribution of the impulses. Such a study suggests that there is a rhythmic excitatory process in nerve, of a fairly constant frequency, which may or may not produce an impulse each cycle. Definite evidence from several sources is now available for the existence of such a process. The role it plays in the regulation of the frequency of conducted impulses will be discussed in the following pages.

The earliest work on the discharge of impulses in single neurons revealed a temporal distribution of impulses that was more or less regular, but not quite periodic. Thus, one of us in 1928, when commenting on the failure of the discharge from a fatigued muscle tension receptor, remarked that "one or more impulses drop out of an otherwise fairly regular series, the impulses becoming more and more scattered."[28] The longer intervals were observed to be approximately equal multiples of the shortest time interval between successive impulses. Adrian observed a similar phenomenon in the discharge of injured mammalian nerve fibers,[13] and such irregular intermittence appears in Pumphrey's[29] records of impulses from taste receptors. More recently, this occasional omission of impulses from an otherwise regular series was observed during the repetitive discharge caused by super-threshold direct current excitation (Erlanger and Blair,[14] Fessard[15]).

This same irregularity in the temporal distribution of impulses is a prominent characteristic in our experiments upon chemically excited axons. This will have been evident in some of the preceding records, but, for the more precise analysis of this phenomenon, additional experiments will be presented. The fibers were excited by removing Ca^{++} from a short length of nerve, by means of sodium citrate, as previously described. The measurements were made on records taken when the nerve was producing impulses at a constant average frequency. Under these circumstances, the temporal distribution may be regular or irregular.

The magnitudes of the time intervals between successive impulses in a certain series obtained in the above manner are plotted in FIGURE 19,

as a function of the number of the impulse in the sequence. Obviously, the intervals are grouped about certain values, which are 6, 12, and 18 milliseconds. These same data are presented in the form of a distribution plot in FIGURE 20, where the number of intervals in a certain range is plotted against the length of the interval. Since the ordinate is a measure of the probability of occurrence of the interval indicated

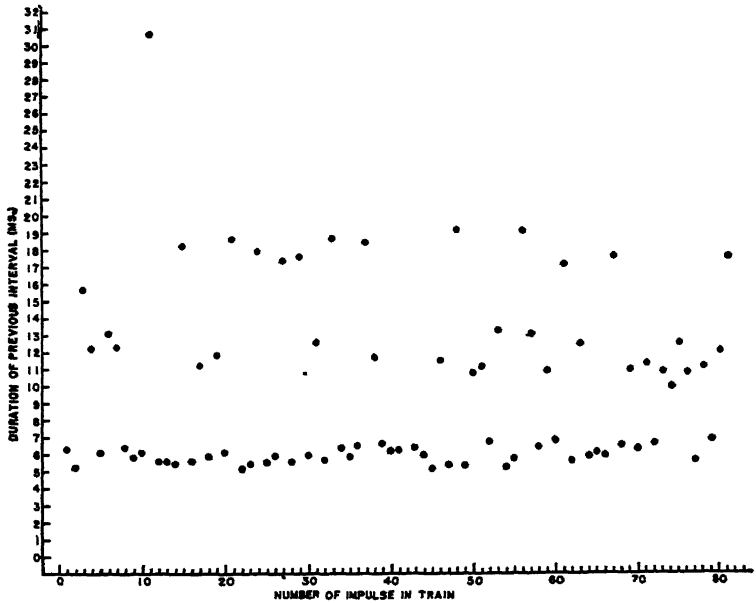


FIGURE 19. The intervals between impulses recorded from a chemically excited single fiber (frog) are approximately integral multiples of a least interval. In this fiber, the least interval was about 6 milliseconds. Stimulation by localized removal of calcium from the axon.

on the axis of abscissae, it is obvious that all intervals are not equally probable. The most probable values are 6, 12, and 18 milliseconds. In all the frog fibers thus far examined, the most probable values of the least interval are in the range of 3–6 milliseconds.

The *aliquot* relations between these most probable values for the fiber just cited suggest, again, that the longer intervals are due to the omission of one or more impulses from an otherwise continuous series. Such omissions could be due to failure of the impulses to be initiated in the chemically excited region, or to blocked conduction between that region and the recording electrodes. If the nerve is stimulated at high frequency by repeated electric shocks, the impulses travel over the nerve fiber and through the treated region. Consequently, there is no reason why each impulse initiated by chemical excitation should not, likewise, be conducted to the recording electrodes. We conclude,

therefore, that the omission of impulses from the series is not due to conduction block. The longer intervals in the records must be due to failure of one or more impulses to be initiated.

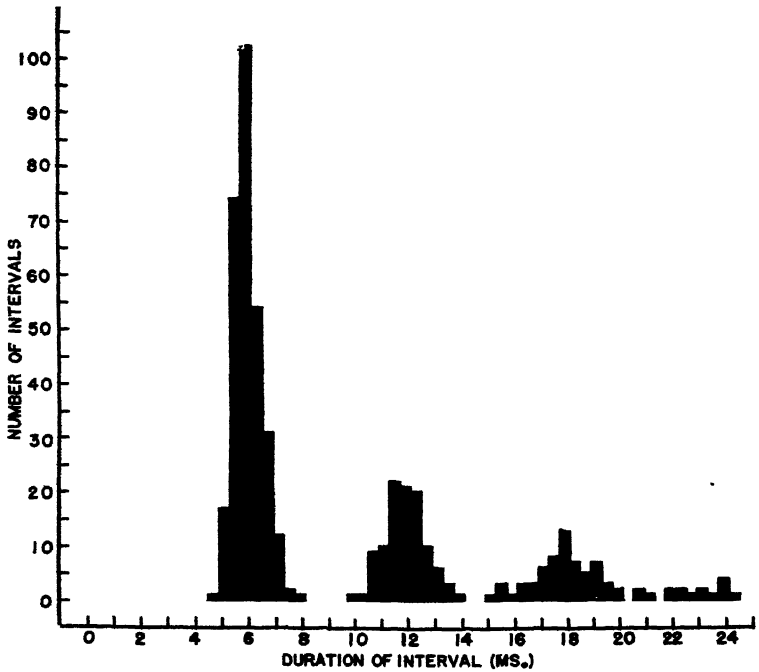


FIGURE 20. Diagram representing the same data as in FIGURE 19, but extended to over 500 successive intervals between impulses. Ordinates are number of intervals having a value in each 0.4 millisecond range. Intervals longer than 24 ms. not shown.

It is difficult to account for these observations, except on the assumption that there is, in this nerve fiber, some rhythmic process, with an average period of 6 milliseconds, that maintains its rhythmic quality independently of the initiation of impulses, once the impulses are started.

Arvanitaki[17] and Hodgkin[31] have presented evidence that an impulse, initiated by electrical stimulation in unmyelinated nerve, develops from a local electrical response which occurs at the site of stimulation. Arvanitaki also showed[17] that this local response, which is elicited by electrical stimulation, may be cyclic in nerves deprived of calcium. Using the giant nerve fiber of the squid, we have studied the development of this local electrical response in nerves excited solely by the removal of calcium. We had two objectives: (1) to see if the local response appears before the initiation of the first impulse in a train, and (2) to ascertain whether a rhythmic local response could be produced chemically, and independently of conducted impulses.

When part of a squid giant fiber is deprived of calcium ions, trains of impulses are initiated in this region and conducted to the recording electrodes. The nerve behaves, in this respect, exactly like the myelinated nerve of a frog, but here conditions are more favorable for recording the difference of potential between the chemically excited region of the fiber and a remote portion. When this is done, it is found that the conducted impulses are preceded by a series of local periodic potential changes of variable amplitude and relatively constant frequency[32] (FIGURE 21). The spacing between adjacent peaks of the



FIGURE 21. Local electrical response, recorded from a chemically excited region of giant axon of squid, is oscillatory and precedes the conducted impulses. The last ten oscillations on the right of the record initiated propagated impulses, which are much larger in amplitude than shown. Stimulation by topical application of isotonic sodium chloride.

local response, just before the conducted impulses appear, is the same as the spacing between the conducted impulses. It is obvious, as Arvanitaki concluded,[17] that the frequency of conducted impulses along the giant axon is determined by the frequency of the local excitatory process.

If relatively little calcium is removed from the nerve, local periodic electric changes may be observed which do not initiate propagated impulses. The local process is an essential part of the excitatory mechanism, but the cyclic changes initiate impulses only when a given cycle is of sufficient magnitude. Furthermore, the frequency of the local process is essentially independent of whether or not a conducted impulse is initiated by each cycle.

FIGURE 22 gives the frequencies of conducted impulses observed in nerves treated with solutions containing different concentrations of Ca^{++}, or with sodium citrate. The frequencies vary from 250 to 400 cycles per sec. This range is comparable to that which is characteristic of the undamped natural frequency of the nerve membrane, as calculated by Cole and Baker[44] from impedance measurements on squid nerve. This

parallelism between the range of frequencies in the local rhythmic response, the fundamental frequency in the trains of chemically initiated impulses, and the undamped natural frequency of the resting nerve membrane, support Cole's suggestion[34] that it is the structural characteristics of the membrane which govern the periodic activity of nerve.

Two frequencies of nerve action have been described in the foregoing discussions of the response of nerve to chemical excitation. One is the average number of impulses conducted along the nerve per second; the other is the fundamental and relatively constant frequency of the excitatory process, which has, in the case of squid nerve, been identified with the local electric response.

Nerve	A 5 mm. Calcium	B No Calcium	C No Calcium. 80 mm. Sodium Citrate
1	260		
2	250		
3	300		
4	250	300	
	260	290	400
5		260	310
6			340
7		260	290
8		300	350
9			240
10			290

Solutions are modified sea water. Calcium and magnesium omitted in B and C. Magnesium omitted in A. KCl concentration and pH are same as in sea water.

FIGURE 22. Values for the fundamental frequency observed in 10 giant axons of the squid (*Loligo pealu*), estimated from frequency of conducted impulses. Stimulation by topical application of indicated isotonic solutions.

The relation between these two frequencies is illustrated by the following experiment, which makes use of the fact that a polarizing current may modify the average frequency of impulses discharged from a calcium-deficient region of nerve. In FIGURE 13, the outward flow of current across the chemically altered nerve membrane caused a transient increase of the average impulse frequency. The distribution plot for the intervals between impulses from the non-polarized nerve is shown in the lower half of FIGURE 23. The intervals between some impulses were 3.2 milliseconds; other impulses recurred at intervals which were about two times this value.

In accordance with the concepts which have been developed in this section, we may say that there was a rhythmic excitatory process, in the chemically modified portion of the nerve, of a fairly constant frequency.

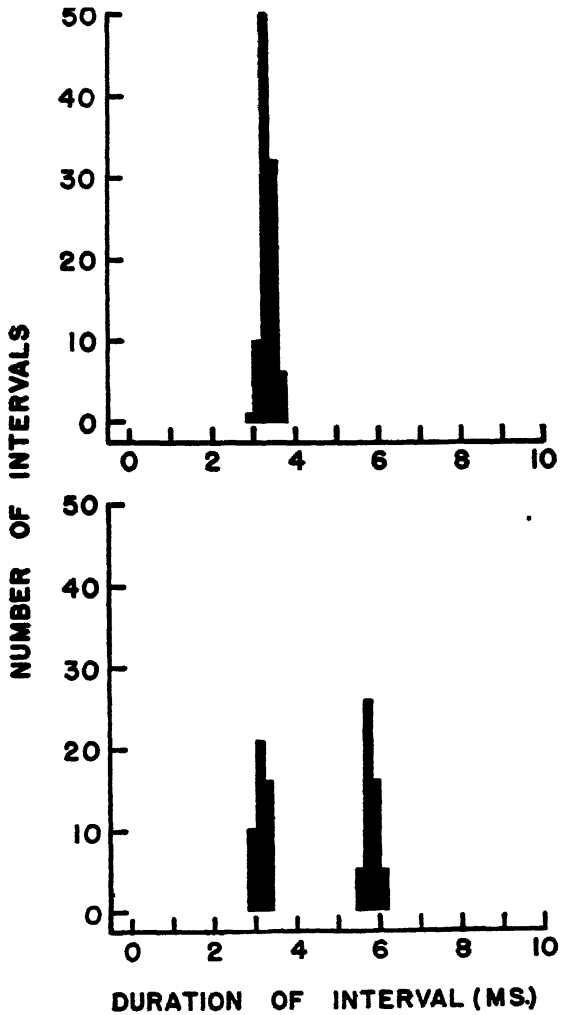


FIGURE 23. The average number of impulses per second in a chemically excited single fiber (frog) can be markedly increased, without much change in the fundamental period.
Lower graph: Frequency of occurrence in a train of impulses of intervals having values indicated on the abscissae. Stimulation by topical application of a solution containing sodium citrate. The average number of conducted impulses per second was about 220.
Upper graph: Similar analysis of a consecutive series of impulses recorded at height of increased response evoked by cathodal polarization of chemically treated region (see FIGURES 12 and 13). The average number per second of conducted impulses was about 305.

Some of these cycles of local change failed to initiate conducted impulses, and, accordingly, some of the intervals between successive impulses were integral multiples of the least time interval of 3.2 milliseconds. The average impulse frequency was 220 per second.

A similar analysis of frequency relations, during the time of increased response caused by the cathodal polarization, was then made. The distribution of intervals between impulses is plotted in the upper part of FIGURE 23. The total number of impulses there considered is the same as in the lower plot. During this time, the most probable period for the excitatory process was only slightly changed, but the probability that impulses would recur at the longer intervals was then practically zero. The action of the polarizing current caused more impulses to be discharged at the basic interval, and this increased the average frequency to 305 per second. In general, changes in this average frequency of impulses, caused by superimposed chemical or physical agents, occur with only slight modification of the most probable least interval. It should be pointed out, however, that, under certain conditions of intense stimulation, this interval can be decreased.

Since the temporal distribution of impulses initiated by chemically excited nerve appears to be regulated by a local rhythmic process, which is determined by the intrinsic characteristics of the nerve, the possibility arises that the same mechanism may govern the discharge of impulses from naturally excited sense organs and motor nerve cells Whether the mechanisms described above do have such a general significance, must wait upon further investigations.

BIBLIOGRAPHY

1. **Misske, B.**
1930. Biochem. Zeit. **219**: 320.
2. **Lipton, S. R.**
1934. Am. J. Physiol. **109**: 457.
3. **Blumenfeldt, E.**
1925. Biochem. Zeit. **156**: 236.
4. **Gerard, R. W.**
1930. Proc. Soc. Exp. Biol. & Med. **27**: 1052.
5. **Brink, F., T. Sjostrand, & D. W. Bronk**
1939. Am. J. Physiol. **126**: P442.
6. **Adrian, E. D.**
1932. The Mechanism of Nervous Action. Univ. of Pennsylvania Press.
7. **Adrian, E. D., & Y. Zotterman**
1926. J. Physiol. **61**: 151.
8. **Adrian, E. D., & D. W. Bronk**
1928. J. Physiol. **66**: 81.
9. **Brink, F., & D. W. Bronk**
1937. Proc. Soc. Exp. Biol. & Med. **37**: 94.
10. **Chang, T. H., M. Shaffer, & R. W. Gerard**
1935. Am. J. Physiol. **111**: 681.
11. **Davies, P. W., & F. Brink**
1941. Am. J. Physiol. **133**: P257.
12. **Bronk, D. W., F. Brink, & P. W. Davies**
1941. Am. J. Physiol. **133**: P224.

13. **Adrian, E. D.**
1930. Proc. Roy. Soc. London B **106**: 596.
14. **Erlanger, J., & E. A. Blair**
1935. Am. J. Physiol. **114**: 328.
15. **Fessard, A.**
1936. Propriétés Rythmiques de la Matière Vivante. I. Nerfs Myelinèses. Hermann & Cie, Paris.
16. **Katz, B.**
1936. J. Physiol. **88**: 239.
17. **Arvanitaki, A.**
1939. Arch. Int. de Physiol. **49**: 209.
18. **Sjostrand, T., F. Brink, & D. W. Bronk**
1938. Proc. Soc. Exp. Biol. & Med. **38**: 918.
19. **Fessard, A.**
1936. Propriétés Rythmiques de la Matière Vivante. II. Nerfs Non Myelinèses. Hermann & Cie, Paris.
20. **Brink, F., T. Sjostrand, & D. W. Bronk**
1938. Am. J. Physiol. **123**: P22.
21. **Cowan, S. L., & W. G. Walter**
1938. J. Physiol. **91**: 101.
22. **Lorente de Nó, R.**
1944. J. Cell. & Comp. Physiol. **24**: 85.
23. **Bullock, T. H., D. Nachmansohn, & M. A. Rothenberg**
1946. J. Neurophysiol. **9**: 9.
24. **Ranson, S. W., & P. R. Billingsley**
1918. J. Comp. Neurol. **29**: 313.
25. **Bronk, D. W.**
1939. Symposium on the Synapse. C. C. Thomas, Baltimore.
26. **Feldberg, W., & A. Vartiainen**
1935. J. Physiol. **83**: 103.
27. **Larrabee, M. G., & D. W. Bronk**
Unpublished.
28. **Bronk, D. W.**
1929. J. Physiol. **67**: 270.
29. **Pumphrey, R. J.**
1935. J. Cell. & Comp. Physiol. **6**: 457.
30. **Erlanger, J.**
1937. Electrical Signs of Nervous Activity. University of Pennsylvania Press, Philadelphia.
31. **Hodgkin, A. L.**
1938. Proc. Roy. Soc. London B **126**: 87.
32. **Brink, F., & D. W. Bronk**
1941. Am. J. Physiol. **133**: P222.
33. **Cole, K. S., & R. F. Baker**
1940–1941. J. Gen. Physiol. **24**: 771.
34. **Cole, K. S.**
1941–1942. J. Gen. Physiol. **25**: 29.

ELECTRICAL CHARACTERISTICS OF ELECTRIC TISSUE

By R. T. Cox, C. W. Coates, and M. Vertner Brown

The Department of Physics, The Johns Hopkins University, Baltimore, Maryland; The New York Aquarium, New York Zoological Society; and the Department of Physics, College of the City of New York.

The group of electric fishes comprises a number of very different varieties, both fresh water and marine. All of them possess special organs capable of producing transient electric discharges, which, in some species, are quite weak, but in others, are powerful enough to give a severe shock. These organs vary widely among the different species in their shape and size and in their position and orientation in the body of the fish. They are alike in having a common unit of structure, the electroplax.

The arrangement of the electroplaxes has its highest geometrical regularity in the electric rays, *Torpedo* and *Narcine*.[1, 2] In the electric organs of these genera, they are piled in columns, an average one of which contains about 400 electroplaxes in *Torpedo marmorata* and perhaps 300 in *Narcine brasiliensis*. Each column extends from the ventral to the dorsal surface of the body. A number of them, side by side, form each of the two electric organs, which lie in the disk-like body of the fish to the right and left of the body cavity, just outside the line of gill slits. In each organ, there are four or five hundred columns in *Torpedo marmorata* and *Narcine brasiliensis*, and about a thousand in *T. occidentalis*. During the discharge, the current traverses each organ in the direction from its ventral to its dorsal face. Thus, the columns of electroplaxes discharge in parallel, while, within each column, the electroplaxes act in series (FIGURE 1).

In *Torpedo* and *Narcine* alike, the electric tissue comprises about one sixth of the whole volume of the fish. In the electric eel, *Electrophorus electricus*, it makes, by contrast, about one half. Organs of such a size must conform, in part, to the shape of the fish, and hence there cannot be so regular an arrangement of the electroplaxes as in the rays. It is customary to distinguish in *Electrophorus* three pairs of organs: the main organs, which extend along the posterior four-fifths of the length of the fish; the much smaller organs of Hunter, which lie under the main organs along their entire length; and the organs of Sachs, which lie

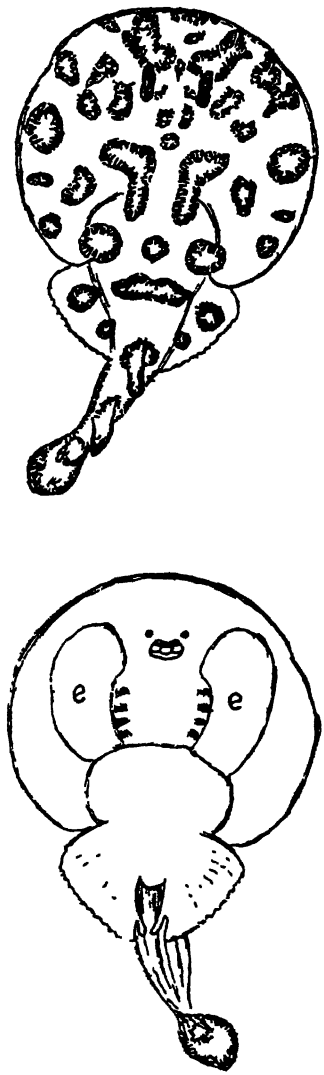
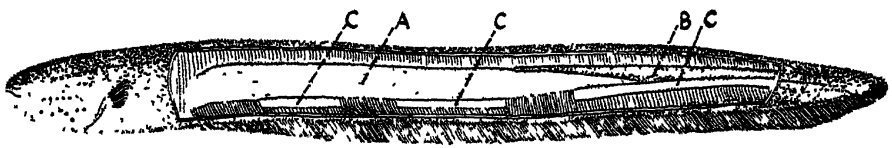


FIGURE 1. Embryo *Narcine brasiliensis*, dorsal and ventral views. Negative faces of electric organs at e, e. (From Zoologica.)

over the main organs in the posterior half of the fish. The organs of Hunter are separated from the main organs only by a thin layer of muscle, and the tissue is identical in adjacent parts of the two pairs of organs. It seems more reasonable to regard the organs of Hunter as parts of the main organs than as a distinct pair (FIGURE 2).

The main organs are of nearly uniform cross-section for some distance from their anterior end, but they taper toward the tail, conforming to the ventral surface of the body and, in the posterior portion, to the under surface of the organs of Sachs. Although this tapering prevents the arrangement of electroplaxes in uniform columns, the series-parallel array already noted in the rays is, nevertheless, clearly discernible. In *Electrophorus*, the axis of polarity is along the length of the fish. Thus, the organs, in comparison with those of the rays, are very much elongated along the line of series connection of the electroplaxes.



Drawing by Ralph Graeter

FIGURE 2. *Electrophorus*, with skin removed to show the electric organs A main organs, B organs of Sachs, overlapping the main organs, C organs of Hunter. (From Bull N Y Zool Soc.)

This contrast between the most highly specialized electric fishes of the fresh-water and marine groups has a significance which seems to have been noticed first by du Bois-Reymond.[1] The combination of a fixed number of electromotive elements to supply power to an external circuit of given resistance is a well-known problem in the theory of electric networks. The solution of the problem shows that maximum power will be delivered to the external circuit by the combination of the elements in a series-parallel array, such that the resistance of the combination is equal to that of the external circuit. Thus, if the external resistance is high, more electromotive elements will be joined in series; if it is low, more will be joined in parallel. Fresh water has a much higher specific resistance than sea water. Consequently, if the condition for maximum external power is equally approximated in the different genera, the organs of the fresh-water fishes will be elongated, and those of the marine fishes will be flattened, along the axis of polarity. Most of the varieties confirm such a generalization. The electric skates, which are marine fishes with weak electric organs elongated in the direction of the axis of polarity, make a rather puzzling exception.

The arrangement of the electroplaxes, by means of which the main organs of *Electrophorus* are accommodated to the tapering body of the fish, is simple and rather interesting.[3, 4] It is best described in terms of a transverse slice just thick enough to contain a single layer of electroplaxes. Near the anterior end of the organs of a fish about 1 meter

long, the thickness of this single electroplax layer is about .01 cm. In fish of this length, the cross-section of the organs near the anterior end has an average area of about 30 cm^2. Hence, the volume of the single electropax layer is 0.3 cm^3. The organs taper caudally, but, as the cross section decreases, there is a compensating increase in the thickness of the layer, and the volume is nearly uniform over most of the length of the organs. The structure is much as if uniform layers were assembled in a long column, and then the column were drawn out thin toward one end, the layers being changed in shape, but not in volume.

These long organs found in *Electrophorus* offer remarkable advantages in the study of the action of electric tissue in the living fish. The series array of electroplax layers is accessible for electrical connection all along its length, rather than only at the ends as in the rays. The variation in structure makes it possible to compare in the same specimen the electrical characteristics of electroplax layers of very different dimensions. Also, *Electrophorus*, which comes to the surface to breathe, can be kept for some time out of water without injury, and the electrical characteristics of its tissue remain constant during an interval in which it can produce a thousand or more electric impulses.

In our observations, the fish is removed from the water and laid in a dry wooden trough. Electrodes made of aluminum strip 1 cm. wide may be placed in any of a number of slots in this trough. These make contact with the skin adjacent to the electric organs and, when they are connected to a cathode-ray oscillograph, it is possible to record photographically the discharge of the part of the organs included between the electrodes (FIGURE 3). The measurements made with the oscillograph are found not to vary appreciably with the area of contact between skin and electrode, provided this area is not less than a few square centimeters. Of course, no appreciable dependence on the choice of a metal for the electrodes is to be expected, since the voltages measured are very much greater than any contact potential differences.

When the electrodes are at the extremities of the main organs of a mature specimen, and the external circuit is open, so that there is no electric current outside the body of the fish, the average peak voltage is about 370 volts.[5] The highest voltage we have measured is 550. There is also a discharge of much lower voltage, which is evidently produced by the organs of Sachs, since it is observed only when some part of these organs lies between the electrodes. In immature specimens, the voltages are smaller. The voltage of the main organs increases with their length, at an average rate of 8 volts per cm., until the organs attain a length of about 50 cm. The organs may ultimately attain

three times this length, but, in any group of longer specimens, the variations in voltage appear to be random.

By measuring the peak voltage between electrodes 5 or 10 cm. apart at different places along the organs, it is possible to compare the voltage per cm. in different parts At the anterior end, where the number of electroplax layers is greatest, the voltage per cm. is also greatest. It decreases caudally, as the electroplax layers thicken. The voltage per electroplax layer is roughly uniform along the organs In four

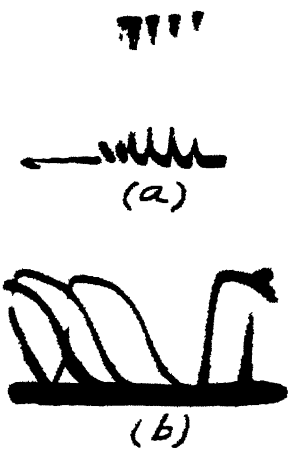


FIGURE 3 Oscillographic traces of the discharge of *Electrophorus*
(a) An impulse from Sachs' organs followed by five impulses from the main organs; sweep period, 50 msec.
(b) Impulses from the main organs, superimposed by successive sweeps; sweep period, 4 msec (From Zoologica)

specimens, values from 0.11 to 0.16 volt were found at the anterior ends of the organs. Somewhat lower values are found in the posterior parts, but, over most of the length, the voltage per electroplax layer is 0.1 volt or more.[3, 4] Values around 0.1 volt per electroplax layer are found also in *Narcine brasiliensis*. In *Narcine*, however, and also in *Torpedo*, the voltage of the organs varies widely with the condition of the fish.

When a conductor is connected between the electrodes, so that the electric tissue produces an external current, the peak voltage is lower than with the external circuit open. If care is taken not to tire the fish, the voltages obtained with a given resistance are reproducible. The resistance R of the external conductor being known, the external current I is found from the measured voltage V, by the relation, $I = V/R$. When conductors of successively lower resistance are employed, the voltage continues to decrease, as the current increases. The results of

such a series of measurements are most conveniently shown by plotting the values of the voltage against the values of the current, one plotted point representing the voltage and current obtained with a given resistance (FIGURE 4). It is found that the points lie near a straight line

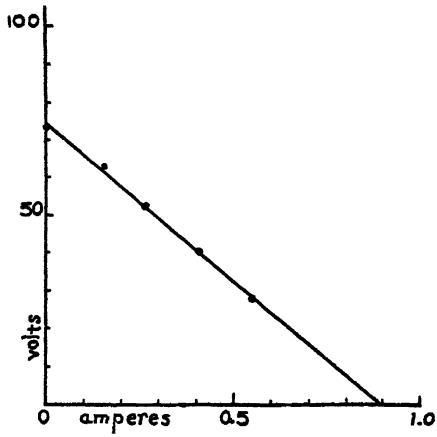


FIGURE 4 Peak voltage vs. peak current in external circuits of different resistance joining electrodes on main organs of *Electrophorus* (From J. Gen. Physiol.)

with both the electric eel and the electric rays.[4] The meaning of this result is that the electric tissue, at least at the peak of the discharge and within the uncertainty of the measurements, can be described electrically in terms of electromotive force and ohmic resistance.

If the external resistance were made negligibly small, the voltage also would be negligible. The corresponding current, estimated by extrapolating the straight line of the graph to zero voltage, is the maximum current of the organs. This maximum current varies from one specimen to another and, in *Electrophorus*, it varies between different parts of the main organs. In an average specimen, around 1 m. in length, it is about 1 amp. at the anterior end of the main organs. Values of about 4 amp. have been found in adult specimens of *Narcine brasiliensis*, and a value of 120 amp. was roughly estimated in a single, very large specimen of *Torpedo occidentalis*.[2, 4, 6]

These great variations are due more to differences in the cross-sections of the organs than to differences in the electrical characteristics of the tissue. The maximum current per unit area of the electroplax layer was found to have values in *Electrophorus* from .02 to .06 amp. per cm.^{2}, the average being about .04. In *Narcine brasiliensis*, the value is about 0.1 and, in the specimen of *Torpedo occidentalis* just mentioned, it was about 0.2 amp. per cm.^{2} (In this calculation and

others to follow, it is implied that the two paired organs discharge simultaneously. In *Narcine* and *Torpedo*, where the organs are far apart, a simple experiment shows this is true. The evidence in respect to *Electrophorus* points to the same conclusion, but it is not certain.)

If opening the external circuit prevented any current in the organs, the voltage measured with the circuit open would be equal to the electromotive force of the part of the organs included between the electrodes. Then, the maximum voltage per electroplax layer would be equal to the electromotive force of the layer, and its quotient by the maximum current per cm.^{2} would be the resistance of 1 cm.^{2} of the layer at the peak of the discharge. However, even with the external circuit open, there must be closed circuits within the body of the fish, during the discharge. Consequently, the voltage per electroplax layer must be somewhat less than the electromotive force of the layer. Its quotient by the maximum current per cm.^{2} is still of some significance as a lower limit for the resistance of 1 cm.^{2} of electroplax layer. It seems likely, also, that this lower limit is not very much less than the actual value. In *Electrophorus,* minimum values thus found for the resistance of 1 cm.^{2} of electroplax layer have varied in different specimens between 2 and 5 ohms. In two specimens of *Narcine brasiliensis,* the values were about 1 ohm.

It is interesting that, in *Electrophorus*, the resistance of unit area of electroplax layer does not increase caudally, although the thickness of the layer increases about ten-fold from the anterior to the posterior end of the main organs. This suggests that the resistance resides principally at boundaries in the electric tissue.

From the observations considered thus far, it is seen that the single electroplax layer in *Electrophorus* has characteristic electric quantities which are roughly uniform, in spite of wide variations in the arrangement of the electroplaxes and in the size of the fish observed. Also, in those cases in which comparison has been possible, it is found that these quantities have the same order of magnitude in *Narcine* and *Torpedo* as in *Electrophorus*.

We have studied, in a number of different specimens of *Electrophorus,* the variation during an impulse of the electrical characteristics of the tissue.[4] The oscillograph was connected, as already described, to electrodes placed 10 cm. apart against the main organs of each specimen studied. Impulses with the external circuit open and closed, through resistances from 400 to 50 ohms, were recorded photographically. With each specimen and each value of the external resistance, measurements were made on a number of oscillographic traces at each

of several short time intervals after the beginning of the impulse. Values of the voltage measured at any one interval, with a given fish and a given external resistance, were then averaged. As with the measurements at the peak of the impulse, the results are conveniently shown by plotting the voltage against the external current. A typical set of measurements is thus shown in FIGURE 5. The points along any one

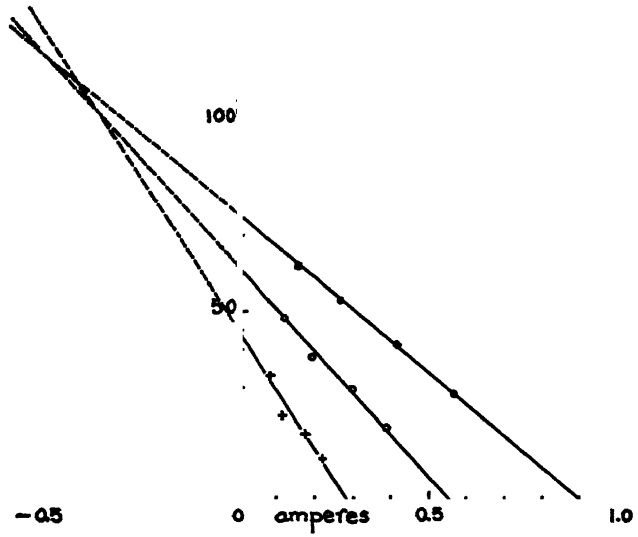


FIGURE 5. Each line shows voltage *vs.* current at one instant, in external circuits of different resistance joining electrodes on main organs of *Electrophorus*. (From J. Gen. Physiol.)

line show values of voltage and current obtained with different external resistances, at the same interval after the beginning of the impulse. The graph farthest to the top and right shows the measurements at the peak and, thus, corresponds to the single graph shown in FIGURE 4. The other two graphs show the measurements at two later instants during the impulse.

Although a straight line cannot be drawn precisely through the plotted points of the measurements at a given interval, we have been unable to detect, in the series of observations, as a whole, any systematic deviation from a linear relation. It appears, therefore, that the tissue can be described electrically in terms of electromotive force and ohmic resistance, not only at the peak of the impulse, but at other times as well.

A simple diagram for such a description is shown in FIGURE 6. In reference to this figure, let E denote the electromotive force of the part of the organs included between the electrodes at p and q, and let r de-

note the internal resistance. The current in whatever circuits are closed within the body of the fish is treated, somewhat arbitrarily, as traversing a single path of resistance, R'.

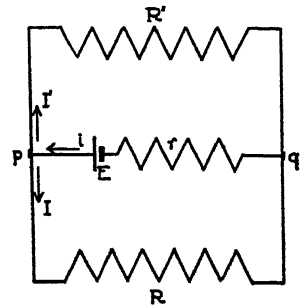


FIGURE 6. Simple diagram for describing the impulse from the main organs of *Electrophorus*. (From J. Gen. Physiol.)

Let i, I, and I' denote the currents in r, R, and R', respectively. Since the current i branches to form the currents I and I', it follows that:

$$i = I + I'.$$

Let V denote the voltage measured by the oscillograph, connected at p and q. This voltage may be reckoned in any of the three branches of the network, and, thus, we obtain three expressions for V, as follows:

$$V = IR, \qquad V = I'R', \qquad V = E - ir.$$

Eliminating i and I', among these four equations, we obtain two expressions for the external current I in terms of the voltage V:

$$I = \frac{V}{R}, \qquad I = \frac{E}{r} - V\left(\frac{1}{r} + \frac{1}{R'}\right).$$

The first of these equations is used to find the current I from the measured voltage V, by means of the known resistance R. The values of I and V are plotted as in FIGURE 5. The other equation is then used to interpret the graph so obtained. If the resistances r and R' are ohmic, they are constants in this equation, which is then a linear relation between I and V, such as is actually found to exist. If, in this equation, we let $V = E$, we find the corresponding value of I to be $-E/R'$. If we suppose that E and R' have the same values at different instants during the discharge, this equation states that the graphs for the different instants, when extrapolated to negative values of I, will intersect at a point. The co-ordinates of this point, moreover, will determine the values of E and R'.

Actually, the lines shown in the figure, which are typical of those obtained from measurements after the peak of the impulse has been attained, do nearly meet in a point. The fact that the graphs of the

measurements at later instants are steeper than the graph of the peak values, indicates a rise in the internal resistance r, after the peak is passed.

These results led us, earlier, to suppose that the electromotive force might be constant throughout the impulse and between impulses as well, the discharge being caused by a transient drop in the resistance from a very high resting value.[4] At that time, we had not succeeded in plotting graphs of voltage and current at instants during the brief interval of rising voltage. Both the steepness of the rising phase of the oscillographic trace and its consequent faintness in the photographs made measurement difficult in this interval. Measurements which we have made more recently have obliged us to reconsider our earlier opinion The graphs of voltage and current obtained from measurements during the interval of rising voltage do not meet at a point. Moreover, even during the interval of falling voltage, we find that deviations, which were formerly within our estimated errors of measurement and which we, therefore, supposed were accidental, appear consistently in the later observations.

The variation in resistance during the interval of falling voltage seems, in any case, well established. It seems probable, also, that the electromotive force is at least approximately constant during this phase. Our immediate object is an estimate of the total electric energy produced in an impulse. Fortunately for this purpose, the time after the attainment of peak voltage is most of the duration of the impulse. Although the changes in the electrical characteristics, during the brief phase of rapidly rising voltage, remain uncertain, the assumptions made about them in the calculation of the energy can be varied widely, without changing the result by more than about 10 per cent.

From the equations already given, it follows that the current i traversing the electric tissue is related to the current I, measured in the external circuit by the equation:

$$i = (1 + R\ R')I.$$

In this equation, R is known, and R' is determined by the intersection of the voltage-current graphs. Thus, the current in the electric tissue is found.

The charge q which passes through the tissue in one impulse is given by:

$$q = \int i dt,$$

where t denotes the time, and the integration is performed over the duration of the impulse. The integration can easily be done graphically.

The charge passing through 1 cm.^{2} of electroplax layer is found by dividing q by the cross-sectional area of the electric organs. (The measurements were made at the anterior end, where the cross-section is nearly uniform.) Since the fish on which the measurements were made were not killed, the cross-section had to be determined indirectly from external measurements. Two methods were employed. In the first method, the girth of the fish was measured. The cross-section of the organs was then estimated by comparison with measurements on dissected fish, on the assumption that the cross-section of the organs bears a constant ratio to the square of the girth. In the second method, a simple mechanical device was employed to trace the outline of the cross-section of the live fish. The area enclosed by this outline was measured, and the cross-section of the organs was taken as 59 per cent of the total area, this percentage having been obtained from measurements on a number of dissected specimens. When both methods were used, the agreement between the results was fairly good, the values determined by the two methods showing a mean deviation of around eight per cent.

(However, the use of only the first method, in another experiment, led to a rather serious error.) The number used then as the ratio of the cross-sectional area of the organs to the square of the girth of the fish was obtained from measurements on only two specimens. Also, the girth of these sections was not measured in the same way as on the live fish, and this led to a further discrepancy, which was increased when the girth was squared. The correction of the resulting error to accord with our new measurements on a larger number of specimens requires that the values of electric energy per gm. and impulse given in the paper referred to should be reduced about 40 per cent. Instead of inferring, as we did in a subsequent paper,[4] that the total electric energy is about equal to that of the breakdown of phosphocreatine and the production of lactic acid, we should now infer that the electric energy is about six tenths of the sum of the energies of these two chemical processes. The correction brings this result into fair agreement with that reported by Nachmansohn, elsewhere in this volume.)

The electromotive force of the part of the organs included between the electrodes was determined by the point of intersection of the voltage-current graphs. This quantity was divided by the distance between the eletrodes, to give the electromotive force per cm. along the column of electroplax layers. The product of the electromotive force per cm., regarded as constant during the impulse, by the charge traversing one cm.^{2} of electroplax layer, is the total electric energy per cm.^{3} produced in one impulse.

The charge passing through the organs, and hence, also, the energy, depend on the resistance of the external circuit, as well as on the electrical characteristics of the tissue. For comparison with the chemical measurements, it is, of course, essential that the external resistance should be the same in the electrical, as in the chemical, experiments. Otherwise, the choice of an external resistance is, within limits, unimportant. The fish on which the electrical measurements have been made form three groups according to size, with average lengths of 67, 103, and 180 cm. We have used an external resistance of 200 ohms with the two groups of smaller length, and of 100 ohms with the other group. These resistances were roughly the same as the internal resistances, averaged over the time of the impulse, of the part of the organs between the electrodes. Consequently, the condition of the experiment approximated the requirement for maximum energy in the external circuit.

The results are summarized in TABLE 1, which shows, for each specimen and for the average of each group, the electromotive force per cm., the charge passing in one impulse through 1 cm.^{2} of electroplax layer, and the total electric energy produced in one impulse in 1 cm.^{3} of electric tissue. With the electrical units employed, the product of the electromotive force per cm. by the charge per cm.^{2} would give the energy per cm.^{3} in microjoules. For more convenient comparison with the chemical energies reported by Nachmansohn, the energies have been given, instead, in microcalories. Since the tissue has nearly unit specific gravity, the energy per cm.^{3} may be taken as the energy per gm., without serious error.

It has already been mentioned that the assumption made in the calculations, that the right and left organs discharge simultaneously, though probable, is still unproved. If the fact should be that the organs discharge separately, then the given values of the charge per cm.^{2} and the energy per cm.^{3} would have to be doubled. The same correction would have to be applied to the values of the chemical energy, and, therefore, the comparison of the electrical and chemical energies made by Nachmansohn is valid in either case.

The individual values of electromotive force per cm. show a mean deviation of 12 per cent from the average of the group. The mean deviation of the charge per cm.^{2} is 15 per cent, and that of the energy per cm.^{3} is 23 per cent. The risk of a serious uncertainty in the averages, beyond that indicated by these deviations, depends on the possibility that the equations used in the computation are seriously in error. The evidence by which they were justified has already been given.

TABLE 1

ELECTRICAL MEASUREMENTS ON *Electrophorus*

	Length of fish (cm.)	Cross-section of electric organs ($cm.^2$)	Electromotive force per cm. (volts)	Charge per $cm.^2$ and impulse (microcoulombs)	Energy per $cm.^3$ and impulse (microcalories)
External resistance, 200 ohms	60	10	20	11	55
	62	12	15	20	72
	65	13	15	18	64
	68	10	19	22	99
	68	16	23	21	118
	70	14	20	20	95
	72	13	21	21	108
	74	13	20	19	89
Ave.	67	13	19	19	88
External resistance, 200 ohms	90	28	16	18	71
	93	33	13	12	35
	96	23	13	14	43
	100	32	18	14	59
	102	38	14	12	40
	104	31	12	10	29
	107	31	18	13	54
	112	36	12	11	32
	121	27	15	16	57
Ave.	103	31	15	13	47
External resistance, 100 ohms	164	67	15	5.0	18
	169	70	12	9.5	27
	186	70	11	8.0	21
	187	67	10	8.7	20
	187	72	13	11.3	35
	188	66	10	10.8	26
Ave.	180	69	12	8.9	24

It seems unlikely that the actual values of the electrical quantities should be much lower than those given in the table. The voltage measured with the external circuit open sets a lower limit to the possible value of the electromotive force. The maximum measured voltage is, on the average for all the specimens, 77 per cent of the value found for the electromotive force. It seems reasonable that the current in circuits closed in the body of the fish should cause a voltage drop of 23 per cent. Similarly, the charge passing through the external resistance, which is obtained directly from simple oscillographic measurements, sets a lower limit to the possible magnitude of the charge tra-

versing the tissue. This lower limit is, on the average, 82 per cent of the computed total charge. It seems unlikely that the external charge is, actually, much nearer than this to the total charge. From these considerations, we should judge it improbable that our assumptions involve a systematic error, whereby the actual values of the energy should be consistently more than about 15 per cent lower than those found by the method we have used.

The question as to how much the actual values may exceed those we find, presents a greater uncertainty. Although, as was just explained, it is difficult to propose an equivalent network for the electric organs which will dissipate much less energy than we allow for in internal currents, there is no difficulty in proposing a network to dissipate more energy. This follows from Thévenin's theorem of electric networks, according to which any electromotive force inferred from external measurements may be regarded, alternatively, as the open-circuit voltage of a concealed network containing a higher electromotive force. On the other hand, our calculations ascribe to the single electroplax layer an electromotive force about as high as any that are found at boundaries in bioelectric phenomena. This suggests that the actual values of the electrical quantities are not very much greater than those we calculate. For the energy, which is the most uncertain of these quantities, twice the calculated value appears to be a safe upper limit.

REFERENCES

1. **Gotch, F.**
1900. The Physiology of Electrical Organs. Textbook of Physiology, Vol. 2. Young J. Pentland, Edinburgh and London; MacMillan, New York. (This article contains references to a number of original sources.)

2. **Cox, R. T., & C. M. Breder, Jr.**
1943. Observations on the electric discharge of *Narcine brasiliensis* (Ölfers) Zoologica **28**: 45.

3. **Cox, R. T., W. A. Rosenblith, J. A. Cutler, R. S. Mathews, & C. W. Coates**
1940. A comparison of some electrical and anatomical characteristics of the electric eel, *Electrophorus electricus* (Linnaeus). Zoologica **25**: 553.

4. **Cox, R. T., C. W. Coates, & M. V. Brown**
1945 Relations between the structure, electrical characteristics, and chemical processes of electric tissue. J. Gen. Physiol. **28**: 187.

5. **Coates, C. W., & R. T. Cox**
1945. A comparison of length and voltage in the electric eel, *Electrophorus electricus* (Linnaeus). Zoologica. **30**: 89.

6. **Coates, C. W., & R. T. Cox**
1942. Observations on the electric discharge of *Torpedo occidentalis*. Zoologica. **27**: 25.

7. **Nachmansohn, D., R. T. Cox, C. W. Coates, & A. L. Machado**
1943. Action potential and enzyme activity in the electric organ of *Electrophorus electricus*. II. Phosphocreatine as energy source of the action potential. J. Neurophysiol. **6**: 383.

SOME BASIC ASPECTS OF THE ACTIVITY OF ELECTRIC PLATES

By Alfred Fessard

Institut Marey, Paris, France

The most conspicuous manifestations of the activity of electric organs are: (a) the electrical discharge itself; (b) variations of the electrical conductivity of the tissue; (c) thermal effects; (d) chemical (metabolic) transformations.

Every individual feature of the total discharge is more or less closely associated with the species of fish examined, with its shape and dimensions, and with the structural arrangement of the physiological units (or *electric plates*) composing the organ. However, on the cellular dimension scale, the behavior of these plates shows a striking unity, and is generally considered as an ordinary bioelectric phenomenon. The classical conceptions of nerve physiology are applicable here, and these, combined with our rapidly progressing knowledge of neuromuscular transmission mechanisms, have been used for explaining the production of electricity in these organs. Here, briefly exposed for each essential manifestation of activity, are the most significant facts in favor of the unitary conception, together with some details concerning our experimental contribution to this problem.

A. The most recent determinations of the electromotive force per plate show, on *Electrophorus electricus*, a fairly uniform value: *viz.*, 0.14 volts (Coates and Cox[11]). This is in agreement with the highest value found by Curtis and Cole[16] for the action potential from the squid giant axon. As for the time course of the discharge, non-typical shapes are often observed (Cox, Coates, and Brown[15]), which may wrongly be considered as representing the elementary process. Our own research in this field has definitely convinced us that these non-typical wave shapes are due to statistical dispersion effects, the causes of which are low velocity at the periphery and differences in length of winding nerve endings.

In our experiments (most of them still unpublished), we used organs of *Torpedo marmorata* and of *Raia undulata*. Our purpose was to record the discharge of a single plate and then to interpret the complex discharges in terms of their components. Columns of electric tissue

were carefully isolated, and transverse slices about 1 millimeter thick cut out with fine scissors. Such fragments contain 10 to 20 plates, many of them damaged by the dissection. The intact ones retain a good excitability for hours. By transverse electrical stimulation at threshold values, it is sometimes possible to record responses which, beyond all doubt, are elementary (PLATE 4A, lower record). They obey the all-or-nothing law and are the conspicuous components of the complex waves obtained with slightly higher intensities. In *Torpedo,* they last 1.75–2.25 msec. (rising phase, 0.5–0.7 msec.); in skates (Auger and Fessard[3]), they are much longer, about 12 msec. (rising phase, 2–2.5 msec.). Now the discharge of a whole column is noticeably of longer duration (PLATE 4A, middle record). In spite of the shortness of the nervous tracts contained in these small pieces of tissue, latencies vary, from less than 1 msec. to more than 4 msec., and dispersion of components lengthens the wave duration to 2 or 3 times its elementary value. The natural discharge starting off the whole organ is hardly longer. The long distance command is transmitted by high velocity fibers of large diameter, and these do not introduce such an important shift in components as do the thin nerve branches at the periphery.

The elementary electrical process is, therefore, comparable to a nerve action potential, never being diphasic, as conduction is absent. Positive after-potentials have never been observed. Mention must be made of some variations in the declining phase, which is sometimes longer than usual, especially in skates. In these cases, the discharge is more like an end-plate potential than an action potential.

B. The analogy between the electric organs and other excitable systems retains its value when the electrical conductivity is studied during the discharge. Using the impedance bridge method, applied by Cole and Curtis[12] to the squid giant axon, we have observed[2] a transient drop of impedance during the activity of *Torpedo* and *Raia* organs. Recently, Cox, Coates, and Brown[15] have obtained indirect evidence of a diminution in ohmic resistance during the discharge of *Electrophorus electricus.*

PLATE 4B shows one of our records obtained with a double beam cathode ray oscillograph. The upper record is the discharge of an isolated column (*Raia*), placed in one branch of an impedance bridge; the lower one shows the reappearance of the 15,000 cycles oscillation feeding the bridge. The two phenomena start simultaneously, instead of showing a shift as in the nerve, a fact easily interpreted as a lack of conduction. The impedance change is slight, less than 10 per cent, but

here the presence of inactive tissue around the plates renders illusory any attempt to introduce quantitative measurements.

C. In observations unfortunately never repeated, Bernstein and Tschermak[10] have shown an initial cooling of the organ, when the discharge was externally derived through a resistance. They interpreted this fact as being inconsistent with a chemical mechanism of the energy supply. Meyerhof[29] criticized this interpretation, without, however, denying the fact, which should be reinvestigated with the more perfected methods now available.

D. Recent investigations into the biochemistry of electric organs proved their metabolism to be quite similar to that of nerve tissue or striated muscle. Cholinesterase has been shown to be very abundant in electric tissue (concerning this significant presence of cholinesterase in electric organs, as at all neuronal surfaces, see Nachmansohn[30]). This fact implies that acetylcholine plays an essential role here, as elsewhere.

It was under the impetus of Nachmansohn that research was undertaken by Feldberg, Fessard, and Nachmansohn to detect the probable presence of acetylcholine in electric organs, and to study the part it plays in the production of the discharge. In these experiments, positive results were obtained in *Torpedo* organs. During stimulation, the ester appears in the perfusate, and arterial injections of micro-doses of acetylcholine produce long-lasting electrical changes.[18, 19]

The next step was when Nachmansohn and his collaborators[31] demonstrated that cholinesterase is localized at the active surfaces in electric organs, its concentration being strictly correlated with the maximum voltage and, consequently, like the electromotive force, with the number of plates per unit length. Now the acetylcholine release is able to start the chain of reactions, beginning with the phosphorylated compounds acting in nerve or muscle metabolism, which we now know to be present, together with the associated enzymatic system, in the electric organs (Baldwin and Needham,[8] Kisch,[25] Nachmansohn *et al.*[32]). The research in this field has now reached a quantitative aspect, and the energy liberated by the breakdown of phosphocreatine and the formation of lactic acid during the discharge can be compared with the electrical energy released (Nachmansohn *et al.*[32]).

In concluding this short survey, we can say that there is a striking convergence of data, allowing the electric plate to be put side-by-side with the nerve and muscle units, from the point of view of their electro-

chemical properties. However, some uncertainties persist as to the particular mode of production of their electrical discharges. We are still ignorant of the way in which acetylcholine may act at an interface to generate electricity. This ignorance is general, but we do not even know, in electric organs, where this active interface lies; and, to assign a definite physiological significance to the plate, we are faced with at least three different views. Although we are far from being able to give a satisfactory answer to these three debated questions, we shall briefly discuss the last two in the light of the experimental evidence we have obtained, up to the present, in our research on the *Torpedo*.

I. PHYSIOLOGICAL SIGNIFICANCE OF THE ELECTRIC PLATE

A. According to a current view, the electric plate is considered as an element of a true effector, and this implies the notion of its physiological individuality. As in the case of the muscle, this special effector would be normally set into activity through a "relay" mechanism, and should show, by direct stimulation, an excitability of its own. This comparison with striated muscle is all the more justified, as both effectors have a common embryonic origin (the *Malapterurus* organ excepted). On the other hand, functional analogies seem to exist between the electrical discharge and the muscular contraction (Marey[28]).

B. However, the regression of all vestiges of striation in the adult stage of the more powerful electric organs; the absence of myosin (replaced by mucin[7]) among the proteins of electrical tissues; and, above all, the simultaneous disappearance of direct and indirect excitability, under different conditions (nerve degeneration, fatigue, cooling) have thrown serious doubts upon the value of the analogy. Some authors have gone so far as to consider the possibility of a purely nervous origin of the discharge. Gotch[21] wrote that "the excitatory electromotive change may be nothing more than the fact that when an excitatory process travels down a nerve, the nerve trunk becomes negative to its terminal cross-section." The maximum value of a nerve action potential is the same as that of an electric plate, and it is suggestive to note that the elementary plate discharge and the single fiber action potential of the nerve commanding the organ have exactly the same duration (PLATE 4A, upper part). Furthermore, the chemical data are far from being opposed to this conception, which tends to reduce the role of the plate to that of a simple support for a richly expanding nervous branching.

C. Nevertheless, as Rosenberg points out in his Review on the subject,[13] it seems unlikely that the plate has only this passive role of support. According to our views, the main difficulty of a purely nervous theory lies in the fact that the orientation of the discharge does not agree with the symmetries offered by the nerve distribution.

An electric organ is sometimes described as an accumulation of modified motor end-plates. This view is more in agreement than any other with the ontogenic facts and with the analogies suggested by histology. There are many nuclei in the plate near the innervated face, as in the sarcoplasmic sole of the striated muscle. Couteaux recently described, at the myoneural junction, a rod-like structure which is strikingly analogous to that long believed to be specific of electric organs.[14]

The existence, now well established, of a localized response at the nerve-muscle junction (end-plate potential, e.p.p.), preceding the muscle fiber propagated impulse, renders the analogy still more evident. This e.p.p., like the discharge, is accompanied by an impedance change that follows the same time-course as that observed in electric organs (Katz[24]): *i.e.*, a non-delayed rising phase and a maximum effect near the inflexion point of the potential variation.

Our experiments on small isolated fragments contribute to show that Hypothesis **A** cannot be retained, as it is really impossible to isolate the plate as a functional effector unit. They are also more in agreement with Hypothesis **C** than with **B**.

a. We thought it useful, at first, to revert to the degeneration test, for the observations mentioned by Garten[21] were not sufficient in number, and the methods for electrical detection have improved since that time. 30 animals were operated on and examined at different intervals after nerve sections on one side. Some survived more than 2 months, and this was sufficient to detect histological signs of alteration in the terminals (Fessard and Pézard[20]). Such signs began to appear on our *Torpedoes* only 5 or 6 weeks after the operation (average temperature 14° C.). Before that, the organs were found excitable, although needing more and more current. Excitability in any form (electrical, chemical, mechanical) totally disappears after about 7 weeks. Osmic acid staining then shows fragmentation of the last branches. Deprived of its terminal innervation, the electric organ is decidedly incapable of activity.

b. No sound conclusion can be deduced from the old results on poisoning by curare. Most of the previous experimenters (namely,

Moreau, Babuchin, Gotch) denied its action, and this was, at first, also our opinion, as 24 hour immersions in 1 per cent curare solutions had no marked effect. In other experiments, we had noted that light cuts made with a razor blade along the longitudinal surface of a column did not interfere with its capacity for delivering good responses. We thought that this treatment might facilitate the penetration of drugs, and we immediately got a positive result (Auger and Fessard[5]). 1 per cent solutions acting during 1–2 hours abolish all excitability. As opposed to the striated muscle, the electric organ becomes inexcitable after curarization.

c. If the electric plate is a real functional unit, its activity must be greatly impaired or completely suppressed by severe mutilation. Isolated columns were divided lengthwise into three narrow strips, each plate being thus fragmented into 3 parts, and severely damaged. This is a complementary situation, compared to that of the nerve degeneration experiments, the plate itself being practically destroyed, but the finer nerve tracts at the endings being only partially damaged. In spite of this drastic treatment, the preparation remains excitable and gives discharges of smaller amplitudes, but of normal shape.

d. One may object to the strict validity of arguments (*a*) and (*b*), as they concern situations in which the plate is modified in some way. For instance, after nerve degeneration, the organ shows some reduction in thickness, and the curare poisoning may have altered the properties of the plate. Results obtained with isolated intact columns may supply indirect, but more satisfactory, evidence.

Using very strong electrical stimuli, we had expected to get a true effector response, as the nerve impulses would arrive during the refractory period of this hypothetical effector unit. Different electrode positions were tried, the results of which we observed from the point of view of threshold, latency, amplitude, and components of complex waves. The results show that any of these parameters (and the variations thereof) depends upon, and can only be explained by, the characteristics of the nerve supply pattern. They appear to be determined by the symmetry of the nerve distribution, not by that of the plates. These are some of our observations:

1. No difference in latency or in the form of the discharge can be observed in supra-maximal longitudinal excitation, whether the current is or is not in the direction of the discharge. No systematic differences in threshold values were found.

2. In the lengthwise stimulation of *Torpedo* columns, the latency can never be reduced to less than a certain minimum (3 to 5 msec.), however strong the current (even if we approach the lethal value). This can be explained by the special distribution of nerves in *Torpedo* columns, if one admits that excitation is always localized at the bending points of the nerve branches. The situation is comparable to that in which Rushton[34] made his observations on excitation of bent nerves. Before entering the plates, the nerve branches run along the edges of the prismatic column, then each axon bends at right angles and, by multiple division at the bending point, sends small transverse twigs to several plates. This point ("*bouquet de Wagner*") is at the same time a Ranvier node, and we assume, as most likely, that it is the most distal one from which excitation can start in longitudinal stimulation.

3. This assumption is confirmed by the fact that the latency is not irreducible. Strong *transverse* stimulations lower it to less than 1 msec. (Auger and Fessard[1]), clearly showing that the long latencies are due to nerve conduction in the plate plane, and not to some elaboration process in the plate itself. When the intensity is lowered, the transverse latency increases regularly, but never exceeds, in normal conditions, the shortest latency observed in longitudinal stimulation. This is perfectly comprehensible, if we localize the excitation at bending points nearer and nearer the "bouquets de Wagner," provided that we adopt the current opinion, according to which the threshold values diminish, the further we are from the nerve extremities.

e. The non-existence of a relay action similar to that of the neuromuscular command is further indicated by the absence of repetitive response when acetylcholine is injected intra-arterially into an isolated organ,[18, 19] although we have shown that this drug exerts, in this case, a marked depolarizing effect.

f. Other drugs were introduced into the interior of the plates by the same technique as described in (*b*) (curare poisoning): eserine (10^{-4}), which lengthens up to more than 4 times the declining phase of the elementary discharge; atropin (10^{-}), which suppresses all excitability in 1–2 hours; curare, which has the same effect as atropin, but in doses ten times larger. During the course of both intoxications, the threshold intensity progressively rises. The duration of the elementary discharge does not change or even become shorter. These data confirm the cholinergic nature of the nerves supplying the organ. They also add supplementary evidence in favor of the similitude between the electric discharge and the end-plate potential (cf. Kuffler[26, 27]).

II. ORIGIN AND LOCALIZATION OF THE ELECTROMOTIVE FORCE

These are largely a matter of speculation, as they now concern the level of molecular organization. Thermodynamic data on one hand, experiments with microelectrodes on the other, would be most useful. Awaiting these, we must content ourselves with discussing the points on which experimental results are available: for instance, the most important problem of plate polarization in the resting state.

a. The classical hypothesis is that of Bernstein.[11] This postulates a permanent superficial polarization of the plate boundaries. The arrival of a nerve impulse results in a local transient annulment of this polarization, in accordance with the general assumption. Now, only one side of each plate is innervated and capable of being depolarized. As, at rest, the potential difference between the two extremities of an electric organ is approximately zero, in spite of the coupling in series of the plates, one must suppose an exact compensation of the electromotive force developed on one side of each plate by that developed on the other side (FIGURE 2, Schema I). We thus arrive at that strange conception of two distant polarized layers, endowed with different properties and yet electrically charged in exactly the same way. Their properties are different, because one is supposed to discharge itself through a sudden internal leak due to collapse, while the other starts discharging without collapse into the external medium. Yet not a sign of a double evolution of potential can be observed in the course of the elementary discharge. On the other hand, these opposite layers are not situated in similar regions, from the point of view of tissue structure and chemical environment. It is very unlikely that they should develop the same electromotive force.

b. Another hypothesis has been recently proposed by Cox, Coates, and Vertner Brown,[13] who assume a constantly present electromotive force, non-compensated by another opposed electromotive force, but hidden by the high resistivity of an interface. This is not conceivable, in our opinion, without capacitive properties by which the interface appears as passively charged (FIGURE 2, Schema II). According to the present concept, "the discharge would be started by a very large and rapid drop in the resistance." We have seen that this drop in resistance really exists, but we cannot conceive of the resting voltage, supposedly present, being lowered, say from 500 volts to 5 millivolts, by the simple interposition of biological membranes, the resistances of which are, at the highest estimate, 1000 ohms/cm.^{2} in the nerve interfaces

(Cole and Hodgkin[13]). Furthermore, electrometric determinations on isolated portions freed from the internal derivations normally present in the intact animal, should reveal much higher resting potentials. This was never observed.

c. It appears to us that the following question should be resolved first: Do polarized layers really exist in the plate, previous to its state of activity? To prove this, it is necessary to communicate in some way with the interior of the plates.

1. We have tried piercing slowly a column from the electric organ of *Torpedo* with a fine metallic electrode and have observed small repetitive discharges, due to a mechanical excitation, which we have shown to be caused by irritation of the nerve twigs encountered.[6] When the electrode is extremely fine, a number of plates may be perforated without being excited and without giving rise to those systematic variations in potential we might expect when passing through one plate to the following.

2. We also took small groups of columns, one of which we slit laterally with the edge of a heated blade. The measurements were started immediately with the ordinary method of opposition, one electrode being placed on the killed region, the other as far as possible from it, on the intact tissue of the same column. The resting potential had to be dissociated from the long-lasting residuals of activity, following the excitation produced by the lesion (cf. Gotch[22]). Three methods were tried: first, allowing the residuals to vanish; second, diminishing their disturbing effect by a transverse arrangement of the electrodes; third, using a degenerated preparation.

The results in these 3 cases are exemplified in FIGURE 1. In FIGURE 1 (2), the A electrode, being a little more dorsal than B, is positive at the start, according to the direction of the discharges. However, it rapidly reverts to its steady potential value, which is negative, relatively, to B. In (3), no initial discharge is present as expected.

In all cases, no value higher than 5 mv. has been obtained for this rest potential. This is 20 to 30 times less than the elementary discharge. We cannot believe that such a discrepancy can be completely due to a shunt effect.

The preceding results, incomplete as they are, throw a serious doubt on the value of the first and second hypotheses. The alleged permanent polarization may not, therefore, exist, at least not at sufficient strength to play the more important part in the discharge. This suggests a third hypothesis that we formulated once,[4] and according to

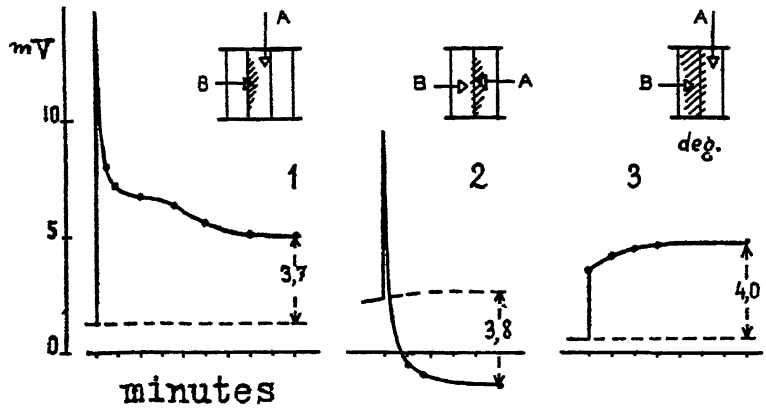


FIGURE 1 Measurements of potential differences (V_A–V_B) immediately following a localized injury (hatched zone)
Torpedo organ, 3 different cases (see text)

which the activity in the electric organ simply consists in the transient appearance of a membrane polarization, rather than in the depolarization of a previously polarized surface. This is the meaning of the schema represented in FIGURE 2, Schema III However, it is difficult

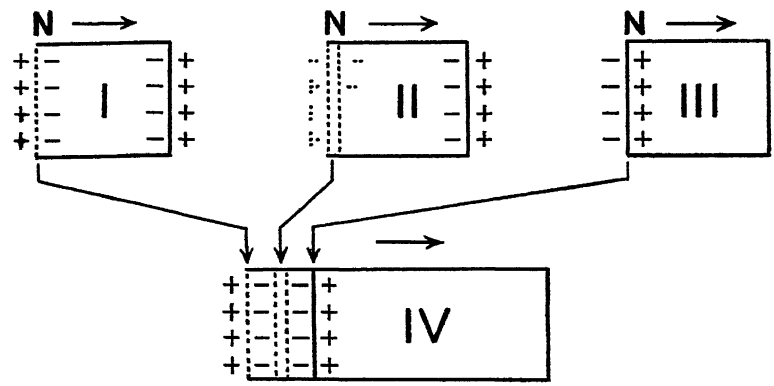


FIGURE 2 Hypothetical schemas proposed to explain the production of an electromotive force in electric organs
N indicates the innervated side of the plate The dotted lines represent what is supposed to collapse during activity
The dotted lines in II are intended to show a passively charged membrane in contrast with other interfaces, which are supposed to be actively polarized by an internal electromotive force
The figure makes it clear that IV may be described as a synthesis of I, II, and III

to believe in the formation of a polarized layer at the moment when the responsible membrane collapses, as is indicated by the drop of its resistance. This schema, like the other two, appears really far too simple.

Finally, the schema that we shall tentatively propose as the most representative of our present knowledge and the most promising as help for future research, is the one shown in FIGURE 2, Schema IV. It is supposed to represent a complex molecular structure of the same type as that used by modern biophysicists in their hypotheses on the constitution of the molecular membranes (Danielli and Davson[17]). Such double layer leaflets are built up with lipoid and protein molecules, and are widespread at the cell surfaces. Similar arrangements have been assumed to be present, for instance, in the rods of the retina, which has well-known electrical properties.

FIGURE 2 makes it clear how this last hypothetical schema may be considered as a synthesis of the other three In effect, it borrows an idea from each of the preceding theories It is like the first, inasmuch as it admits the presence of two opposing polarized layers, of which only one can be neutralized. It borrows from the second the idea that a high resistivity layer exists (probably made of oriented lipoid chains in the intermediate region), which collapses during the discharge; and it is in agreement with the third, in admitting that the electromotive force is strictly localized at the innervated face of the plate, where it becomes apparent during the short period when one of the layers is depolarized.

BIBLIOGRAPHY

1. **Auger, D., & A. Fessard**
1938. C. R. Soc. Biol. **128**: 1067.
2. **Auger, D., & A. Fessard**
1939. Ibid. **131**: 765.
3. **Auger, D., & A. Fessard**
1939. Ann. de Physiol. **15**: 261.
4. **Auger, D., & A. Fessard**
1939. Livro de Homenagem Prof. A. e M. Ozorio de Almeida. 25. Rio de Janeiro, Brazil.
5. **Auger, D., & A. Fessard**
1941. C. R. Soc. Biol. **135**: 76.
6. **Auger, D., & A. Fessard**
1929. Ibid. **102**: 582.
7. **Bailey, K.**
1939. Biochem. J. **33**: 255.
8. **Baldwin, E., & D. Needham**
1937. Proc. Roy. Soc. London B **122**: 197.
9. **Bernstein, J.**
1912. Elektrobiologie. 118.
10. **Bernstein, J., & A. Tschermak**
1906. Pflueger's Arch. **112**: 439.
11. **Coates, C., & R. Cox**
1945. Zoologica **30**: 89.

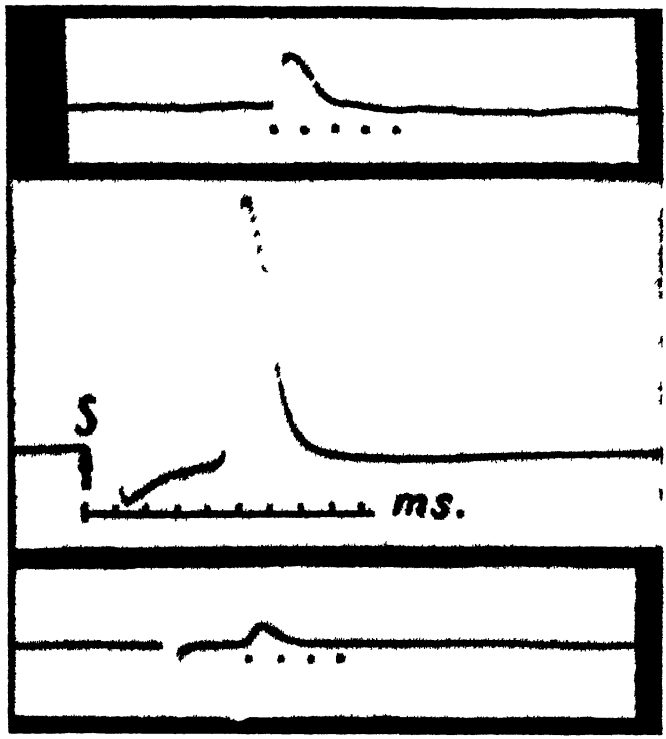
12. **Cole, K., & H. Curtis**
1939. J. Gen. Physiol. **22**: 649.
13. **Cole, K., & A. Hodgkin**
1939. J. Gen. Physiol. **22**: 671.
14. **Couteaux, R.**
1945. C. R. Soc. Biol. **139**: 641.
15. **Cox, R., C. Coates, & V. Brown**
1945. J. Gen. Physiol. **28**: 187.
16. **Curtis, H., & K. Cole**
1942. J. Cell. & Comp. Physiol. **19**: 135.
17. **Danielli, J., & H. Davson**
1943. The Permeability of Natural Membranes. Cambridge.
18. **Feldberg, W., A. Fessard, & D. Nachmansohn**
1939. J. Physiol. **97**: 2.
19. **Feldberg, W., & A. Fessard**
1942. J. Physiol. **101**: 200.
20. **Fessard, A., & A. Pézard**
1940. C. R. Soc. Biol. **134**: 525.
21. **Garten, S.**
1910. Handb. d. vergl. Physiol. 105.
22. **Gotch, F.**
1887. Phil. Trans. B. 487.
23. **Gotch, F.**
1888. Phil. Trans. B. 329.
24. **Katz, B.**
1942. J. Neurophysiol. **5**: 169
25. **Kisch, B.**
1930. Biochem. Z. **225**: 183.
26. **Kuffler, S.**
1942. J. Neurophysiol. **5**: 18.
27. **Kuffler, S.**
1943. Ibid. **6**: 99.
28. **Marey, E.**
1887. Travaux du Laboratoire de l' Institut Marey **3**: 1.
29. **Meyerhof, O.**
1926. Thermodynamik des Lebensprozesses. Handb. d. Physik **11**: 254.
30. **Nachmansohn, D.**
1945. Vitamins and Hormones **3**: 337.
31. **Nachmansohn, D., R. Cox, C. Coates, & A. Machado**
1942. J. Neurophysiol. **5**: 499.
32. **Nachmansohn, D., R. Cox, C. Coates, & A. Machado**
1943. Ibid. **6**: 383.
33. **Rosenberg, H.**
1928. Handb. der norm. u. pathol. Physiol. **8**: 876.
34. **Rushton, W.**
1927. J. Physiol. **63**: 357.

PLATE 4

PLATE 4

A Middle record response of an isolated column (*Torpedo* organ) to an electrical stimulus directed along its main axis
Lower response of a small fragment of tissue, transverse stimulation, threshold intensity
Upper single fiber action potential (electric organ nerve)
Time scales in milliseconds

B Double beam oscillograph record
Above discharge of a fragment of electric tissue (*Raia*)
Below impedance change test with an alternating current of 15 000 cycles
The whole activity wave lasts 12 milliseconds



A



B

FESSARD ACTIVITY OF ELECTRIC PLATES

PHYSIOLOGICAL FUNCTION FROM THE STANDPOINT OF ENZYME CHEMISTRY

By D. E. Green

Departments of Medicine and Biochemistry, College of Physicians and Surgeons, Columbia University, New York, N. Y.

It is a curious fact that, although there is general recognition and agreement that the cell is a chemical system, none the less the full implications of this truism have yet to be appreciated in some fields of physiological investigation. Perhaps the explanation is to be found in the preoccupation of biochemistry, until very recently, with problems of the structure of cellular constituents and with their estimation. Classical biochemistry represented to the physiologist the extension of histology to the chemical field. The study of what we may call chemical morphology was hardly calculated to attract physiologists or to arouse their interest in the chemical basis of physiological function. However, the interest of biochemistry has been shifting gradually from the purely structural problems to the dynamic chemical events of the cell. Our present knowledge of the chemical mechanisms of the cell has grown sufficiently for it to be ignored no longer by those who are concerned with the study of physiological function.

We may conceive of the cell as a chemical factory in which literally thousands of chemical reactions take place, cheek by jowl, without mutual interference. Some of these reactions are concerned in the synthesis of structural components of the cell and others in providing chemical energy for carrying on the activities of the cell. Practically without exception, these reactions do not proceed spontaneously. They require the presence of protein catalysts, which we call enzymes. Each enzyme is distinct, chemically, from all the others, and is uniquely specialized for its particular catalysis. If this picture of the cell is correct, then it follows that all dynamic activities including physiological function must be reducible to terms of enzyme chemistry. In other words, physiological function and enzyme chemistry are two sides of the same coin I hope, in the short time at my disposal, to marshal the available evidence which justifies this interpretation.

In the syndromes of avitaminosis, we observe profound morphological and physiological abnormalities. The recognizable signs of each of the vitamin deficiencies are too well known to be discussed here. The point

I want to make, however, is that these abnormalities which cover the entire gamut of physiological dysfunction can be explained completely in terms of certain enzyme systems. Vitamin B_1, B_2, B_6, and the P-P factor have all been shown to be the prosthetic groups of certain enzyme systems. When these vitamins are not available in the diet, the active enzymes cannot be formed in the cell. In consequence of the failure of these enzyme systems to function properly, an abnormal physiological situation develops, which, if uncorrected, will lead to death. Which organ first registers the effect of a particular deficiency is determined by the amount of the reserves of enzymes containing the vitamin and by the relative importance of this set of enzymes in the economy of the organ. Thus, in B_1 deficiency in the pigeon, the brain is the first organ to register disturbed function, presumably because there are no reserves of this vitamin in the brain and because the active enzyme formed by the vitamin plays a key role in the metabolism of brain. One may well raise the point that, if, as in the avitaminosis, the causal link between the physiological disturbance and the effect on enzyme systems is unquestioned, then surely there is a good case for assuming the same link between the normal physiology and enzyme systems.

Woolley[1] has introduced the use of anti-vitamin reagents which, by virtue of their structural resemblance to the vitamins, are able to prevent the vitamins either from being incorporated into enzymes or from functioning as prosthetic groups. These anti-vitamins produce, in a relatively short period, the same syndromes which arise from depriving the diet of an animal of a particular vitamin. Not only are these anti-vitamins valuable for speeding up the process of avitaminosis, but the profound pharmacological effects which they induce permit correlations between the action of reagents on certain enzyme systems and the pharmacological consequences. In other words, the anti-vitamins have focussed attention on the fact that the effects of certain, if not all, pharmacological agents can be explained completely in terms of effects on enzyme systems.[2, 3] In recent years, a rich literature has grown up to deal with this correlation. There are now at least 14 instances in which the pharmacological effects of certain reagents can be explained in terms of a specific effect on an enzyme system. Thus, iodoacetic acid, which induces muscle rigor, has been found to inhibit, in minute concentration, the triosephosphoric dehydrogenase which catalyzes an essential step in lactic acid formation. This paralysis of the triosephosphoric dehydrogenase accounts for all the pharmacological effects produced by iodoacetic acid. Fluoroacetic acid, the highly

toxic agent discovered in Chemical Warfare Research, has been shown to inhibit the enzyme systems involved in the metabolism of acetic acid. The classical pharmacological reagents, strychnine, eserine, and prostigmine, have been shown to exert their effects exclusively by virtue of their paralysis of cholinesterase. The effects of cyanide on cytochrome oxidase, fluoride on enolase, and chlorine on the triosephosphoric dehydrogenase, are other examples of this correlation. Perhaps even more unexpected, has been the identification of various toxins with enzymes. Thus, spreading factor, the agent which facilitates the rapid diffusion of injected substances through dermal tissue, has been shown to be identical with hyaluronidase, a mucolytic splitting enzyme. The hemolytic principle of *Clostridium welchii* toxin and that of snake venom have been shown to be lecithinases, and the hemolytic effects are completely explicable in terms of their ability to weaken the lipoid membrane of the red blood cell by hydrolysis of the lecithin contained therein. During the war, some English workers, led by McFarlane,[4] identified one of the toxins produced by the gas gangrene organism as collagenase, a proteolytic enzyme which dissolves the connective tissue sheath of muscle. The action of this enzyme explains the pulping of muscle observable in advanced cases of gas gangrene poisoning. At the present moment, it would be premature to assume that all specific pharmacological agents which work at high dilutions are active by virtue of their effects on enzyme systems. On the other hand, it is pertinent to point out that no other principle of mechanism has been established for any pharmacological agent which has been studied. Apart from the dictates of caution, there is no good reason not to anticipate that, eventually, all effects of specific pharmacological agents will be reducible to terms of enzyme chemistry.

The study of endocrines has always been one of the most active fields of physiological investigation, and it is of interest to inquire to what extent hormones can be related to enzyme phenomena. Until quite recently, hormones were held up as notable exceptions to the rule that substances which act at high dilutions must be enzymes or parts of enzymes, or must specifically affect some enzyme system. Some recent research, however, fails to confirm the hormones as exceptions to the enzyme-trace substance thesis. No doubt, everyone is aware of the epoch-making discovery of Cori and his group,[5] that one of the hormones of the anterior pituitary inhibits the action of hexokinase, and that, in turn, this inhibition is released by insulin. We have here a clear blueprint for the way in which hormone antagonism can be effected. A key enzyme system which controls some metabolic process

can be regulated by a set of hormones, one of which inhibits, while the other releases the inhibition. All students of endocrinology have long been aware that hormones regulate metabolic processes, and it is not surprising to find in one instance, at any rate, that the regulation operates at the level of the enzyme systems. Houssay and his colleagues in the Argentine have presented cogent evidence that renin, a kidney hormone, is a type of proteolytic enzyme which hydrolyzes one of the plasma proteins to form a pressor substance. In this instance, the hormone regulates metabolic processes by actually assuming an enzymatic role.

In still another direction, there has been confirmation that enzymic phenomena underlie essential physiological processes. The brilliant work of Beadle[6] and his school have made it abundantly clear that the regulation of growth and development by the hereditary units of the cell, *viz.*, the genes, is exercised through control of enzyme systems. They have shown that each gene determines the synthesis, probably, of a single enzyme. Whereas some of the hormones regulate metabolic reactions by slowing up or speeding up an enzyme reaction, genes regulate by determining the synthesis of an enzyme Remarkably little is known of the mechanisms by which enzymes are synthesized, but it would appear that, whatever the mechanism, the genic material will be implicated.

The mere recognition that enzymic phenomena underlie physiological function is, of course, only the first step in the biochemical analysis. Obviously, the exercise of physiological function requires a source of energy, and the energy must arise in enzyme-catalyzed reactions But how is the energy converted into the manifold forms required by the cell? How is chemical energy converted into mechanical energy of contraction or electrical energy of nervous conduction (to mention two examples)? There are no transforming elements in the cell, such as the storage battery. Until recently, this problem of energy conversions was shrouded in deepest fog, but some light has managed to penetrate. It now appears, from the work of various laboratories, that the contraction of the myosin molecule may be coupled with the enzymatic hydrolysis of adenosine triphosphate. The contraction of muscle is now visualized as the integration of the single contractions of myosin molecules arranged in linear series. The picture is, of course, very crude, and probably will be modified by further research. However, if the basic facts are correct, then we have a blueprint for visualizing energy transfers at the enzyme level. Adenosine triphosphate represents a readily tapped supply of the chemical energy gen-

erated by the process of glycolysis. Because of the close proximity of adenosine triphosphatase and myosin, some of the chemical energy of hydrolysis is absorbed by the myosin molecule, which then undergoes simultaneous contraction. In other words, myosin is acting as a kind of transformer element for the conversion of chemical energy to mechanical energy.

We have to consider the possibility that, just as myosin is specialized for muscular contraction, chlorophyll or visual purple for photochemical reactions, and hemoglobin for oxygen transfer, so there may be one or more proteins in nerve specialized for the reactions which underlie the propagation of a nerve impulse. The knowledge that acetylcholine and adrenaline are the chemical agents involved in nerve conduction, is merely the introduction to the problem of mechanism Undoubtedly, these substances react with special proteins. It is the transformations which these special proteins then undergo that is the basis of the phenomenon of nerve transmission.

REFERENCES

1. **Woolley, D. W.**
1945. Science. **100**: 579.
2. **Green, D. E.**
1946. Currents in Biochemical Research: 149 Interscience Publishers, Inc., New York.
3. **Green, D. E.**
1941 Advances in Enzymology. **1**: 177. Interscience Publishers, Inc., New York.
4. **McFarlane, R. G., & J. D. MacLennan**
1945. Lancet **2**: 328.
5. **Price, W. H., Carl F. Cori, & Sidney P. Colowick**
1945. J. Biol. Chem. **160**: 633.
6. **Beadle, G. W.**
1946. Currents in Biochemical Research: 1. Interscience Publishers, Inc. New York.

CHOLINESTERASE

By Oscar Bodansky*

Medical Division, Chemical Warfare Service, Edgewood Arsenal, Maryland†

On the basis of his studies on the pharmacology of acetylcholine, in 1914, Dale stated: "In the blood at body temperatures it seems not improbable that an esterase contributes to the removal of the active ester from circulation."[1] In 1926, Loewi and Navratil observed that acetylcholine, as well as "vagus substance," was rendered inactive by incubation with heart extract.[2] However, such inactivation did not occur after the heart extract had been heated or subjected to ultraviolet irradiation. These observations by Loewi and Navratil inaugurated the study of the enzyme, cholinesterase.

A heat-labile substance which is capable of hydrolyzing acetylcholine is found very widely distributed in the organs and fluids of the body. It cannot be assumed that, apart from this common property of hydrolyzing acetylcholine, the other properties of this enzyme are the same in all these tissues. Our present discussion should most fittingly concern itself with the properties of this enzyme as found in nervous tissue, and should determine the extent to which these properties play a role in nerve activity. However, most of the data available for discussion describe chiefly the properties of cholinesterase found in serum, red cells, and, to a lesser extent, in the whole brain. The extent to which these data apply to the properties of cholinesterase, at synapses and in other nerve tissue, should be carefully evaluated.

RELATION BETWEEN SUBSTRATE CONCENTRATION AND REACTION VELOCITY

We shall first turn our attention to the relation between the rate of action of cholinesterase and the concentration of the substrate, acetylcholine. Examination of the data shows that two types of relationships hold. The first type appears to follow the Michaelis-Menten formulation:[3]

$$\frac{v}{V_{max}} = \frac{S}{S + K_s} \qquad (1)$$

where v is the reaction velocity at substrate concentration S, V_{max} is the maximum reaction velocity occurring at infinite substrate con-

* Lt. Colonel, M. C., A. U. S.
† Present address: Dept of Pharmacology, Cornell University Medical College, New York, N. Y.

centration, and K_s is the dissociation constant of the intermediate enzyme-substrate complex.

The second type of relation obtaining among some cholinesterases appears to be one in which inhibition of reaction velocity occurs at higher substrate concentrations. This relation has been found to hold for some other enzymes, such as lipases, catalase, oxygenase.[4] The reactions between substrate and enzyme may be formulated as follows:

$$\begin{aligned} E + S &\rightleftarrows ES \qquad (active) \\ ES + (n-1)S &\rightleftarrows ES_n \qquad (inactive) \\ ES &\rightarrow E + P. \end{aligned}$$

The relation between substrate and reaction velocity has been developed by Haldane[5] for the reaction where $n = 2$, as follows:

$$v = \frac{V_{max} \cdot S}{S + K_s + \frac{S^n}{K_2}} \tag{2}$$

where K_2 is the dissociation constant of the inactive enzyme-substrate compound $\frac{[ES][S]^{n-1}}{[ES_n]}$. When the velocity is plotted against the logarithm of the substrate concentration, a bell-shaped curve is obtained. The values for K_s, K_2 may be obtained by transformation of the above equations, according to the method of Lineweaver and Burk.[6] Since there are very few and incomplete data in the literature for cholinesterases which follow this type of substrate-reaction velocity relationship, we shall not attempt such an evaluation.

In TABLE 1, are listed the results of investigations on the relationship between concentration of the substrate, acetylcholine, and the velocity of cholinesterase action. It may be seen that the cholinesterases present in human and dog serum, and in the cat superior cervical ganglion, show increasing rates of reaction, with increasing substrate concentrations, which become asymptotic to a maximal rate at infinite substrate concentration. These values of the dissociation constants (molar concentration at which one-half maximal reaction velocity occurs) range, in general, from 1 to 1.7×10^{-5}. In contrast, the cholinesterases from the red cells of man, sheep, horse, and ox, and from the brain of the mouse and dog, do not show increasing reaction velocities with increasing substrate concentration. According to Mendel and Rudney,[13] and to Alles and Hawes,[11] they show optimal activity at about 1×10^{-5}M; at concentrations higher than this, the reaction velocities decrease.

On the other hand, values for the optimal concentration obtained by Zeller and Bissegger,[12] and by Nachmansohn and Rothenberg,[14] are

TABLE 1

RELATION BETWEEN REACTION VELOCITY OF CHOLINESTERASE ACTION AND ACETYLCHOLINE CONCENTRATION

In accord with Michaelis-Menten formula

Enzyme source	Investigator	Dissociation constant
Dog serum	Goldstein[6]	1.25×10^{-3}
Dog serum	Eadie[7]	1.7×10^{-3}
Dog serum	Wright & Sabine[5]	0.26×10^{-3}
Human serum	Glick[9]	1.1×10^{-3}
Human serum	Wright & Sabine[5]	1.2×10^{-3}
Cat superior cervical ganglion	Glick[10]	1×10^{-3}

Inhibition at higher substrate concentrations (bell-shaped curve)

Enzyme source	Investigator	Optimal concentration (M)
Red cell: human	Alles & Hawes[11]	0.25×10^{-6}
Red cell: human	Zeller & Bissegger[12]	4.4×10^{-3}
Red cell: human, sheep, horse, ox	Mendel & Rudney[13]	$< 1.5 \times 10^{-4}$
Red cell: human	Nachmansohn & Rothenberg[14]	ca. 5×10^{-2}
Brain: mouse, dog	Mendel & Rudney[13]	$< 1.5 \times 10^{-4}$
Brain: human	Zeller & Bissegger[12]	4.1×10^{-3}
Brain: mouse	Nachmansohn & Rothenberg[14]	ca. 6×10^{-3}
Brain areas: ox	Nachmansohn & Rothenberg[11]	ca. 8×10^{-3}

much higher. As Mendel and Rudney[13] have pointed out, and as will be seen later, the concentration of electrolyte influences the optimal concentration of substrate.

The relation existing between reaction velocity and substrate concentration of acetylcholine may be of physiologic significance. Glick[10] found that, under *optimal substrate conditions*, pH 7.4 and 38°, 0.10 γ acetylcholine chloride was split per second, per milligram of cat superior cervical ganglion. Brown and Feldberg[16] found that the concentration of acetylcholine in the superior cervical ganglion of the cat was

22.5γ/gm. of tissue. If conditions of maximal substrate concentration are assumed to exist, 0.225 seconds would be required to split the acetylcholine. If the assumption is made, that lower substrate concentrations exist, then a longer period would be required. Brown and Feldberg also found that 0.1γ acetylcholine was liberated from a ganglion weighing 12.9 mg. and perfused with eserinized Lockes' solution, during the first 5 minutes of preganglionic stimulation at 17 per second. According to Glick, if the substrate concentration were maximal, 78 milliseconds would be necessary to hydrolyze this amount, or 1.5 milliseconds for splitting the acetylcholine liberated by one nerve impulse. These times will be longer, if lower substrate and enzyme concentrations are assumed. The time for hydrolysis, at the minimum rate, was 8 seconds, and localization of enzyme and substrate within the ganglion cell would have to be assumed, to explain enzymatic destruction of acetylcholine liberated by nerve impulses within the span of the refractory period of 2 milliseconds.

EFFECT OF pH AND ELECTROLYTES ON CHOLINESTERASE ACTIVITY

The effect of pH on enzyme activity is, of course, well known. Bernheim & Bernheim,[17] employing a pharmacological procedure, found a pH optimum of 8.4 for the serum and brain cholinesterases of some lower animals. Glick[9, 18] found practically this same value, 8.4 to 8.5 for the cholinesterases of human serum, pig's gastric mucosa, horse serum, and cat brain. Calculations from the shape of the pH activity curve, for these various cholinesterases, show that the activity at 7.4 is about 65 to 70 per cent that of the activity at optimal pH. A somewhat lower optimal pH, 7.5–8.0, has been reported for red cell cholinesterase.[13]

The effect of various ions on the activity of cholinesterase has been studied by several groups of investigators. The activating effect of Ca^{++} and Mg^{++} is well established and, except perhaps for the magnitudes of the degree of activation, appears independent of the source of the enzyme. Thus, Nachmansohn[19] found that 4×10^{-5} M Ca^{++} increased the activity of dialyzed *Torpedo* electric organ cholinesterase 5-fold, and 4×10^{-5} Mg^{++} increased the activity about 8- to 9-fold. Massart and Du Fait[20] found that horse serum, which lost 40 per cent of its cholinesterase activity, on dialysis, regained its normal activity in the presence of 2×10^{-3} M Mg^{++} or Ca^{++}. Mn^{++} has been found to activate considerably both dialyzed *Torpedo* electric organ cholin-

esterase and dialyzed horse serum cholinesterase. Ba^{++} activates *Torpedo* cholinesterase, but not dialyzed horse serum cholinesterase.

There has been considerable disagreement, regarding the activating effects of Na^+ and K^+. But here, as in other respects, these disagreements are resolved, if it is recognized that the studies have been carried out on cholinesterases from various sources. Thus, 1×10^{-1} M K^+ inhibits Mendel and Rudney's purified horse serum cholinesterase 40 per cent at a substrate concentration of .0015 M acetylcholine, 15 per cent at .03 M.[13] Glick[21] found, at .02 M substrate concentration, that Na^+ and K^+ did not have any regular effect on horse serum cholinesterase activity, but both increased rabbit serum cholinesterase activity 25 per cent at 0.3 M Na^+ or K^+, 35 per cent at 0.5 M, and 40 per cent at 1.0 M. Nachmansohn[19] reported that both Na^+ and K^+ activated *Torpedo* cholinesterase.

There appears to be a relationship between the effect of Na^+ and K^+ and the effect of substrate concentration on reaction velocity. Thus, in human serum where the substrate-reaction velocity relationship follows the Michaelis-Menten formulation, 0.16 M NaCl inhibited the cholinesterase about 10 to 20 per cent, at substrate concentrations ranging from 1×10^{-6} to 4×10^{-4}. On the other hand, among red cell cholinesterases, in which reaction velocity decreases at higher substrate concentrations (Haldane formulation), 0.16 M NaCl, inhibited at low substrate concentrations, activated at substrate concentrations greater than about $1 \times 10^{-5.7}$ M. Alles' and Hawes' results[11] show, and Mendel and Rudney[15] have emphasized this point particularly, that, for those cholinesterases which follow the Haldane formulation for the relationship between substrate concentration and reaction velocity, the presence of Na^+ or K^+ not only increases the value of the optimal reaction rate, but also shifts the optimum to higher substrate concentrations.

So far as anions are concerned, cyanide has been found to have no effect on horse or human serum cholinesterase. Oxalate, fluoride, and citrate inhibit dialyzed horse serum cholinesterase to the extent of 30 per cent at 0.002 M concentrations of these ions, 60 per cent at 0.02 M.[20]

INHIBITION OF CHOLINESTERASE ACTIVITY

The inhibiting effects of various compounds on enzyme activity have been generally formulated in terms of an inactive, but reversible, enzyme inhibitor complex.[3, 5] If the inhibition is non-competitive, then

$$\frac{v'}{v} = \frac{K_I}{K_I + I}. \qquad (3)$$

Where v' is the velocity in the presence of the inhibitor, v in its absence, I is the concentration of the inhibitor, and K_I is the dissociation constant of the enzyme-inhibitor complex.

If the inhibition is competitive, then the following expression holds:

$$\frac{v'}{v} = \frac{S+K_s}{S+K_s\left(1+\frac{I}{K_I}\right)} \tag{4}$$

where the terms have the meanings as described in EQUATIONS 1 and 3. K_I may be calculated from the value of the dissociation constant, K'_s, in the presence of a constant concentration of inhibitor, as follows:

$$K_I = \frac{I}{\frac{K'_s}{K_s} - 1} \tag{4'}$$

The Michaelis-Menten derivation is based on the assumption that the concentration of enzyme centers is constant and, as compared with the concentration of any substance with which it could combine, so small that it may be neglected. Recently, Straus and Goldstein,[22] elaborating upon the ideas of Easson and Stedman,[23] have submitted a more general formulation for the effect of an inhibitor, which takes into account those possibilities in which the concentration of enzyme centers may not be negligible. In the presence of a large excess of substrate,

$$I = \frac{K_I\, i}{1-i} + iE \tag{5}$$

total inhibitor — free inhibitor — bound inhibitor

where I is the concentration of total inhibitor, combined and free; i is the fraction of total enzyme combined with inhibitor; E is the concentration of total enzyme; and K_I is the dissociation constant of the enzyme-inhibitor complex. When the above equation is divided by K_I, the following expression is obtained:

$$I' = \frac{i}{1-i} + iE' \tag{6}$$

where $I' = I\ K_I$ and $E' = E/K_I$.

Simplifications of these equations are possible, under conditions where E' is very small, or very large, so that the other term on the right hand side of the equation may be neglected.

The implication of EQUATION 6 is that the degree of inhibition depends upon the value of E', namely, the ratio of the concentration of enzyme centers to the dissociation constant of the complex. This may be illustrated by taking values from a theoretical plot by Straus and

Goldstein. Thus, the log values of I', at $i = 0.5$ (50 per cent inhibition) are as follows, for various values of E':

E'	$\log I'$
0.1	0.00
10	0.79
100	1.71
1000	2.70

It may, thus, be seen that dilution of the enzyme influences the E' value, and hence, the extent of inhibition.

The values of K_I and E may be calculated from the experimental determination of the inhibition, i, at various concentrations, I, of inhibitor. EQUATION 5 may be transposed, as follows:

$$\frac{I}{i} = K_I \times \frac{1}{1 - i} + E \qquad (7)$$

A plot of $\frac{I}{i}$ against $\frac{1}{1 - i}$ should, therefore, yield a straight line, the slope of which would be K_I and the intercept on the Y-axis would be E.

Goldstein[6] has also developed an expression for competitive equilibrium between enzyme, substrate, and inhibitor:

$$I' = \underbrace{\left[(S' - aE_{S'})\left(\frac{1 - a}{a}\right) - 1\right]}_{\text{free}} + \underbrace{\left[1 - a\left(1 + \frac{1}{S' - aE_{S'}}\right)\right]E_{I'}}_{\text{combined}} \qquad (8)$$

total

Where $I' = I\ K_I$, $S' = S\ K_S$, $E'_I = E/K_I$,

$E_S' = E, K_S$, and $a = 1 - i$ or fractional activity of the enzyme.

Another type of derivation is possible, if E_f, the amount of free enzyme, is considered negligible in comparison with the amount of enzyme combined with inhibitor and substrate. Then:

$$I' = \underbrace{(S' - aE_{S'})\left(\frac{1 - a}{a}\right)}_{\text{free}} + \underbrace{(1 - a)E_{I'}}_{\text{combined.}} \qquad (9)$$

total

Various simplifications of EQUATIONS 8 or 9 are possible, depending upon whether we assume $E_{I'}$ or $E_{S'}$ to be small enough to be neglected, or so large that other terms not involving them become negligible.

The investigations on *in vitro* inhibition may now be summarized. In TABLE 2, are shown those results which have been formulated, in terms of the dissociation constants of an inhibitor-enzyme complex, in accordance with the equations already discussed. Several points of interest may be noted. First, the dissociation constants of the enzyme complexes of physostigmine and prostigmine are of a low order of magnitude, about 10^{-7} to 10^{-8}, as compared with the dissociation constants of the cholinesterase-morphine derivative complexes, 10^{-3} to 10^{-4}. The

values of Eadie[7] seem extraordinarily low, but these dissociation constants have been calculated on the assumption that one molecule of enzyme combines with two of inhibitor. Goldstein[6] has discussed this point and has claimed that Eadie's method of determining reaction velocity involved a constant error. The titration of released acetic acid, for twenty minutes, immediately following the addition of enzyme

TABLE 2

DISSOCIATION CONSTANTS OF CHOLINESTERASE-INHIBITOR COMPLEXES

Dissociation constants calculated by method of Michaelis-Menten, except for Straus & Goldstein values. Values for K_I at 37–38°.

Inhibitor	Source of cholinesterase	Investigator	Dissociation constant, K_I
Physostigmine	Horse serum, 22%	Straus & Goldstein[22]	3.7×10^{-8}
Physostigmine	Dog serum, 4.54%	Goldstein[6]	3.11×10^{-6}
Physostigmine	Dog serum	Eadie[7]	3×10^{-14} (n=2)
Prostigmine	Dog serum	Eadie[7]	2×10^{-14} (n=2)
Morphine	Human serum	Wright & Sabine[8]	8.1×10^{-4}
Morphine	Dog serum	Eadie[7]	14.6×10^{-4}
Dilaudid	Human serum	Wright & Sabine[8]	1.2×10^{-3}
Codeine	Human serum	Wright & Sabine[8]	4.2×10^{-4}
Desomorphine	Human serum	Wright & Sabine[8]	1.8×10^{-4}

to a substrate-inhibitor mixture, corresponded to a stage of the reaction where equilibrium had not yet been established, and where the reaction velocities were higher than at equilibrium. This discrepancy was greatest for concentrations of inhibitor producing moderate inhibitions.

A second point of interest is the difference in the dissociation constants of the complexes of the same inhibitor with cholinesterases of different sources. Although the difference is not marked, it may be seen that the constants for the morphine complex are 8.1×10^{-4} for human serum cholinesterase, and 14.6×10^{-4} for dog serum cholinesterase. The occurrence, among different cholinesterases, of differing sensitivities to inhibition by the same compound, will be discussed further.

In TABLE 3, are shown the concentrations of cholinesterase inhibitors giving 50 per cent of the uninhibited cholinesterase activity. It will be readily recognized that these values are equivalent to the values of the dissociation constants, calculated by the Michaelis-Menten expression for non-competitive equilibrium. This table illustrates the order of inhibition of different compounds. It may be seen that human serum, rabbit serum, and human muscle cholinesterases are less sensitive to inhibition by physostigmine than is the horse serum cholin-

TABLE 3

CONCENTRATIONS OF CHOLINESTERASE INHIBITORS GIVING 50 PER CENT OF UNINHIBITED VELOCITY

Inhibitor	Source of Cholinesterase	Investigator	Concentration of Inhibitor
			M
Physostigmine	Horse serum, 4.5%	Collier & Allen[24]	2.5×10^{-8}
Physostigmine	Horse serum, 22.2%	Ellis, Plachte, & Straus[25]	5×10^{-8}
Physostigmine	Horse serum, 11.1%	Ellis, Plachte, & Straus[25]	2.5×10^{-8}
Physostigmine	Human serum, 2.5%	Mazur & Bodansky[26]	2×10^{-7}
Physostigmine	Rabbit serum, 12.5%	Mazur & Bodansky[26]	1.2×10^{-7}
Physostigmine	Human muscle	Mazur & Bodansky[26]	8×10^{-7}
Methylene Blue	Horse serum	Collier & Allen[24]	1.2×10^{-6}
Methylene Blue	Horse serum	Massart & DuFait[27]	6×10^{-6}
Acriflavine	Horse serum	Collier & Allen[24]	6.6×10^{-6}
Phenothiazone	Horse serum	Collier & Allen[24]	6.7×10^{-5}
Thionin	Horse serum	Massart & DuFait[27]	11×10^{-5}
Thiamin	Horse, rat serum	Glick[28]	1.8×10^{-3}
Isopropyl antipyrin	Guinea pig serum	Zeller[29]	2×10^{-3}
Isopropyl antipyrin	Horse serum	Zeller[29]	0.93×10^{-3}

esterase. The basic dyes, methylene blue, acriflavine, phenothiazone, and thionin, are somewhat less potent inhibitors than physostigmine. Next, in order of potency of inhibition, are thiamin and the antipyrines. TABLE 4 shows that several antipyrines at a concentration of 3×10^{-3} M inhibit human serum cholinesterase from 30 to 80 per cent. Sulfonamides at this concentration inhibit only slightly. Acid dyes (*e.g.*, congo red) and p-phenylenediamine have been reported to have no inhibiting action at 2×10^{-4} M.[27]

TABLE 4

INHIBITION OF HUMAN SERUM CHOLINESTERASE ACTIVITY BY ANTIPYRINES AND SULFONAMIDES

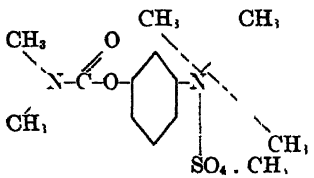
(Zeller)

Inhibitor	Concentration *M.*	Inhibition per cent
Antipyrine	3×10^{-3}	51
Aminoantipyrine	3×10^{-3}	30
Dimethylaminoantipyrine	3×10^{-3}	38
Isopropylantipyrine	3×10^{-3}	80
Sulfanilamide	2.5×10^{-3}	−1
Irgamid	2.5×10^{-3}	13
Sulfathiazole	2.5×10^{-3}	18
Benzyl sulfanilamide	2.5×10^{-3}	9

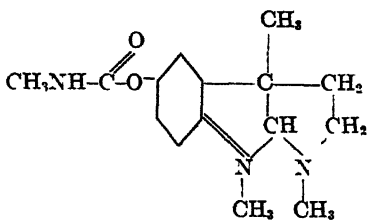
It is, of course, always of interest to attempt a correlation between chemical structure of inhibitors and enzyme action. In general, compounds which resemble the substrate, chemically, inhibit enzyme action, presumably by combining with the same chemical groupings on the enzyme molecule. However, it must be stressed that other chemical groups on the inhibitor molecule may influence this combination. The evaluation of this influence requires considerable experimentation. In the case of cholinesterase, relatively few data are available.

Acetylcholine is a quaternary ammonium compound. A survey of the inhibitors which we have discussed shows that, although no strict correlation can be drawn between structure and degree of inhibition, the closer the inhibitor comes to possessing a completely alkylated nitrogen grouping, the greater is its inhibition. Thus, prostigmine, one of the most effective inhibitors, has a quaternary ammonium grouping. Physostigmine has two tertiary amine groupings. Methylene blue, which is also a potent inhibitor, may be considered to have a quaternary ammonium grouping. Thionin, which is the unalkylated congener of methylene blue, is a less powerful inhibitor; about 20–100 times as great a concentration is needed to produce 50 per cent inhibition.

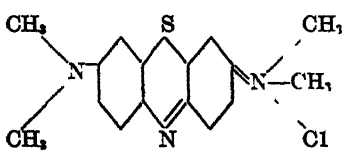
That an alkylated amino grouping is important in inhibition, seems to be generally true. Thus, the pyrazolone derivatives (antipyrines), morphine, caffeine, percaine, are all moderate inhibitors. Unalkylated amino groupings tend to make a compound a poor inhibitor: for example, diphenylamine does not inhibit at 2×10^{-4} M. Acid groupings (COOH, SO_3, H, OH), apparently tend to negate the inhibiting power of alkylated amino groups. Examples of this occur in the sulfonamides and the acid dyes.



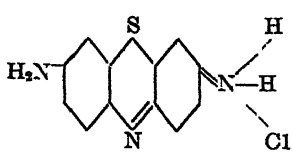
prostigmine



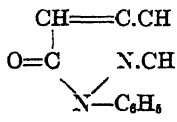
physostigmine



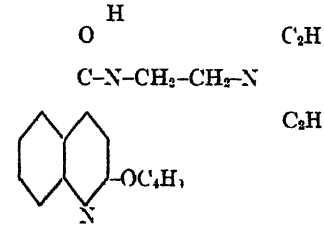
methylene blue



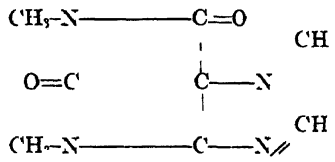
thionin



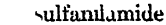
antipyrine



percaine



caffeine



sulfanilamide

It must be noted, however, that the relationship between the degree of inhibition and the chemical structure of the inhibitors also depends upon the source of the cholinesterase. TABLE 5 shows that percaine, irgamid, and isopropyl antipyrine inhibit human serum cholinesterase much more markedly than human red cell or human brain cholinesterase. In contrast, morphine inhibits these three cholinesterases to about the same extent, and caffeine inhibits human serum cholinesterase only slightly, as compared with its effect on human red cell and human brain cholinesterase. Nachmansohn and Schneeman[11] have observed that caffeine inhibits nervous tissue cholinesterases much more markedly than those of horse serum and guinea pig pancreas. In contrast, the inhibitions of these latter cholinesterases by quinine, quinidine, cocaine, and lobeline are more marked than those of nervous tissue.

TABLE 5

EFFECT OF DRUGS ON INHIBITION OF VARIOUS CHOLINESTERASES
(Zeller and Bissegger[12])

Inhibitor	Concentration M	Per cent inhibition		
		Human serum	Human red cells	Human brain
Percaine	6×10^{-3}	94	25	12
Irgamid	6×10^{-3}	46	4	3
Isopropyl antipyrin	6×10^{-3}	81	26	..
Morphine	6×10^{-3}	66	76	66
Caffeine	6×10^{-3}	4	42	40

In addition to the inhibitors which we have discussed above, a number of other substances have been reported to possess an inhibiting ac-

tion: vitamin K, ether, chloral, and some hormone preparations, such as estrone, progestin, testosterone, etc.[31, 32] Observations on these inhibitions are not yet detailed, and their exact significance is not yet defined. It should be recognized that, in general, any enzyme is subject to *in vitro* inhibition by a great number of compounds, particularly at relatively high concentration. Such inhibitions probably do reflect an interaction of the inhibitor with the enzyme molecule, but the *in vivo* significance of such inhibitions is very often questionable.

There is evidence to indicate that the inhibitors which we have already described form reversible enzyme-inhibitor complexes. Matthes,[33] for example, first showed that dialysis of a mixture of cholinesterase and physostigmine resulted in the restoration of the enzyme activity. Similarly, dilution of cholinesterase physostigmine mixtures results in a relative increase of cholinesterase activity, presumably as the result of the dissociation of the inactive enzyme-inhibitor complex.[22, 26] Zeller has shown that the inhibition of cholinesterase by pyrazolons and sulfonamides is similarly reversible, by dialysis of the corresponding enzyme-inhibitor complexes.[29]

IN VITRO AND *IN VIVO* INHIBITION BY DIISOPROPYL-FLUOROPHOSPHATE (DFP)

We should now like to present a description of the properties of a compound, typical of an entire group, which, in contrast to the inhibitors we have described above, forms a combination with cholinesterase which it has, so far, not been found possible to reverse. This compound is diisopropyl-fluorophosphate. It is one of a group of alkyl fluorophosphates first described by Lange and Krueger.[34] During the war, it was regarded as a potential chemical warfare agent, and its properties were first investigated by British workers. Adrian, McCombie, B. A. Kilby, and M. Kilby[35, 16] noted the similarity between the cholinergic effects of the fluorophosphates and those of physostigmine. Mackworth[17] found that incubation of the alkyl fluorophosphates with horse serum cholinesterase resulted in the inactivation of the enzyme, and that dialysis of the fluorophosphate-cholinesterase mixture did not result in any restoration of cholinesterase activity. Our interest in the mechanism of the anticholinesterase action was first aroused when we noted that, upon exposure of men to very low concentrations of this agent, the serum cholinesterase was very markedly reduced to 2 to 5 per cent of the pre-exposure value, in spite of the fact that there were only slight or doubtful systemic symptoms. The

in vitro and *in vivo* inactivation of cholinesterase by DFP has been studied by Mazur and Bodansky,[26] and the results of these studies will now be briefly described.

In Vitro Inhibition of Cholinesterase Activity by Diisopropyl-Fluorophosphate and by Physostigmine. The inhibition of the activities of serum, red cell, muscle, and brain cholinesterases of the rabbit, monkey, and man were determined, at various concentrations of DFP and physostigmine. In order to obtain a general measure of the extent of inhibition of the different cholinesterases, the relative velocities were plotted against the negative logarithm of the molar concentration of DFP or physostigmine. The negative log molar concentration at which 50 per cent inhibition occurred, was termed the pC_I value. These values for various enzyme preparations are shown in TABLE 6.

TABLE 6

SENSITIVITY OF VARIOUS CHOLINESTERASES TO INHIBITION BY DIISOPROPYL-FLUOROPHOSPHATE AND PHYSOSTIGMINE

The values are expressed in terms of the negative log of the concentration of inhibitor required to produce a 50 per cent inhibition of cholinesterase activity (pC_I).

	Serum	Red cells	Muscle	Brain
		Diisopropyl-fluorophosphate		
Rabbit	4.1	5.2		5.5
Monkey (*M. rhesus*)	7.8	5.5		5.5
Human	7.7	5.4	5.4	6.0
Horse	8.3			
Horse* (purified)	8.1			
		Physostigmine		
Rabbit	5.9			
Human	6.7		6.1	

* Horse serum cholinesterase (purified), prepared by Drs. Northrop and Kunitz, according to directions of Mendel and Rudney

It may be seen that, of the various serum cholinesterases studied, that of the rabbit was least sensitive to inhibition by DFP. Thus, a negative log molar concentration of 4.1 of DFP was necessary for 50 per cent inhibition of rabbit serum cholinesterase, whereas concentrations of about one ten thousandth as much (negative log molar values of 7.7 to 8.3) gave 50 per cent inhibition of monkey, human, and horse serum cholinesterase activity. The various red cell cholinesterases

showed approximately the same degree of sensitivity toward inhibition by DFP (pC_I values of 5.2 to 5.5), whereas human brain cholinesterase ($pC_I = 6.0$) was somewhat more sensitive than monkey or rabbit brain cholinesterase ($pC_I = 5.5$) to inhibition by DFP. Purified cholinesterase of the electric eel gave a pC_I value of 4.1.

It is of interest to compare the sensitivities of the various tissue cholinesterases from one species to inhibition by DFP. Rabbit brain and red cell cholinesterases showed a greater sensitivity than did serum cholinesterase; brain showed the greatest sensitivity. Thus, a 50 per cent inhibition of serum cholinesterase activity occurred at a negative log of the molar concentration of DFP of 4.1, whereas the same degree of red cell and brain cholinesterase inhibition occurred at values of 5.2 and 5.5, respectively. Monkey serum cholinesterase was much more sensitive to inhibition by DFP than red cell or brain cholinesterase. Human serum cholinesterase was much more sensitive to inhibition by DFP than human red cell or brain cholinesterase. This picture is similar to that found in the monkey, and is in marked contrast to that observed in the rabbit.

TABLE 6 also shows the sensitivities of rabbit serum, human serum, and muscle cholinesterases to inhibition by physostigmine. It may be seen that rabbit serum cholinesterase was more sensitive to inhibition by physostigmine ($pC_I = 5.9$) than by DFP ($pC_I = 4.1$), whereas the reverse was true with human serum cholinesterase.

The possibility existed that the differences in sensitivity of different cholinesterases to inhibition by DFP were due to materials, other than the enzymes themselves, present in the preparations. TABLE 6 shows that a purified horse serum cholinesterase preparation had, within experimental error, the same pC_I value as horse serum itself. Heat-inactivated extracts of one tissue, added to a tissue possessing cholinesterase activity, did not alter the sensitivity of the latter to inhibition by DFP. Thus, human brain extract was heated to destroy its cholinesterase activity, and added to human serum. The pC_I value for the mixture was 7.7, the same as that found for human serum cholinesterase itself.

In Vivo Inhibition of Cholinesterase Activity by DFP. The extent to which various cholinesterases are inhibited *in vivo*, after administration of DFP, may be considered to depend, not only on the sensitivity of the particular tissue cholinesterase to inhibition by DFP, but also on the localization and, hence, of the concentration of DFP in the tissue. In rabbits exposed to DFP vapor, severe muscular tremors and death

occurred at the higher exposures; at lower exposures, no symptoms except miosis occurred. In most instances, the decrease in red blood cell cholinesterase activity was greater than that in serum cholinesterase activity. It will be recalled that *in vitro* rabbit red cell cholinesterase is more sensitive than serum cholinesterase to inhibition by DFP. When rhesus monkeys were exposed to DFP, the plasma cholinesterase activity was decreased to only 1 to 5 per cent of normal, at almost all exposures, whereas the red cell cholinesterase activity showed only slight decreases at the lower exposures. This marked difference in the extent of decrease paralleled the considerable *in vitro* difference between the sensitivities of monkey red cell and serum cholinesterases to inhibition by DFP.

The effect of intravenously injected DFP on cholinesterase activity, in the rabbit, is shown in TABLE 7 It can be seen that, in most in-

TABLE 7

THE EFFECT OF INTRAVENOUSLY INJECTED DIISOPROPYL-FLUOROPHOSPHATE ON RABBIT PLASMA, RED CELLS, AND BRAIN CHOLINESTERASE ACTIVITY *In Vivo*

Animal number	Dose mg. per kg.	Blood sample time min.	Relative cholinesterase activity*			Remarks
			Plasma per cent	Red cells per cent	Brain per cent	
327	3.0	4	0	0	0	Died immediately
326	0.3	23	15	7	12	Muscle tremors
328	0.3	26	5	0	5	Muscle tremors
491	0.1	26	54	41	59	No symptoms
492	0.1	25	37	29	57	No symptoms
330	0.05	27	60	19	74	No symptoms
331	0.05	43	51	29	73	No symptoms

* These values are per cent of the pre-exposure values.

stances, the red cell cholinesterase activity is more markedly reduced than that of the serum cholinesterase. This difference corresponds to the *in vitro* difference in sensitivities. The brain cholinesterase activities were reduced to about the same extent as the serum cholinesterase activities at the lower doses, and slightly less at the higher doses. This finding did not correspond to the *in vitro* sensitivity of brain cholinesterase to inhibition by DFP. The brain cholinesterase activities, corresponding to the appearance of symptoms, were less than about 60 per cent of normal.

The effect of intravenous injections of DFP in the monkey are shown in TABLE 8. A dose of 0.3 mg. per kg. was fatal in 10 minutes. The serum, red blood cell, and brain cholinesterases were reduced to very

low or zero levels of activity. Essentially the same results were obtained at doses of 0.2 and 0.25 mg. per kg., except that the survival period was longer. At a dose of 0.1 mg. per kg., the animal survived, although the serum and red cell cholinesterase activities were very low. At 0.02 mg. per kg., the serum cholinesterase activity was reduced to a very low level, 2 per cent of normal, whereas the brain cholinesterase activity was decreased only slightly, to 78 per cent of normal. This

TABLE 8

THE EFFECT OF INTRAVENOUSLY INJECTED DIISOPROPYL-FLUOROPHOSPHATE ON MONKEY PLASMA, RED CELLS, AND BRAIN CHOLINESTERASE ACTIVITY *In Vivo*

Animal number	Dose mg. per. kg.	Blood sample time min.	Relative cholinesterase activity *Per cent of pre-injection value*			Remarks
			Plasma	Red cells	Brain	
4	0.3	10	0	2	0	Died in 10 minutes
6	0.25	27	0	1	0	Died in 33 minutes
3	0.2	2 hrs.	0	0	0	Severe symptoms, sacrificed
1	0.1	24	0	2	..	No symptoms
2	0.1	3¾ hrs.	0	10	0	Muscle tremors, diarrhea
5	0.02	60	2	14	78	No symptoms

marked difference between the decreases of serum and brain cholinesterase was similar to that obtained at low exposures of monkeys to DFP vapor, and parallels strikingly the *in vitro* difference in sensitivity between the monkey serum and brain cholinesterases to inhibition by DFP. As will be seen presently, these findings are quite similar to those obtained in man. Although the red blood cell cholinesterase was of about the same order of *in vitro* sensitivity as the brain cholinesterase, it was reduced *in vivo,* at the very low dose, to a greater degree.

In the monkey, symptoms were absent, even when the serum cholinesterase levels were zero. In most instances, the occurrence of severe symptoms or death was associated with zero levels of brain cholinesterase activity. After exposure to, or injection of, DFP, the serum cholinesterase returned very slowly to normal. In the few instances in which this return was followed, about 50 per cent of the normal serum cholinesterase activity was regained in about 7 days.

Men were also exposed to DFP vapor. One group of 7 men was exposed to a concentration of 19 micrograms per liter, for 8⅔ minutes;

a second group of 6, to a concentration of 27.1 micrograms per liter, for 9 minutes; and 2 men, to 28.8 micrograms per liter, for 10 minutes, 40 seconds, and 27 micrograms per liter for 6 minutes, 20 seconds, respectively. The symptoms were extremely mild. All of the men showed miosis and most of them complained of a slight feeling of tightness in the chest, lasting for several hours. The following symptoms were observed occasionally: increased nasal secretion, nausea, salivation.

TABLE 9

EFFECT OF INHALATION OF DIISOPROPYL-FLUOROPHOSPHATE VAPOR ON SERUM CHOLINESTERASE ACTIVITY *In Vivo* IN MAN

Subject	Concentration of DFP vapor	Duration of exposure	Relative cholinesterase activity at various intervals after exposure*									
			5-30 min.	Days								
				1	2	3	4	6	8	10	15	17
	γ per l.	min.										
J.H.	19	8.7	4	13			29		58			71
R.L.	19	8.7	3	7			28		53			77
W.B.	19	8.7	2	14			3		55			72
M.G.	19	8.7	3	12			31		50			69
J.P.M.	19	8.7	1	7			30		52			68

* These values are per cent of the pre-exposure values.

TABLE 9 shows the decreases in serum cholinesterase activity, immediately after exposure, and the rate of return of the activity to normal. It may be seen that, immediately after gassing, the serum cholinesterase activity decreased to about 1 to 5 per cent of the pre-exposure value. The rate of return to normal was very slow. On the average, the activity returned to about 30 per cent, in four days; to about 50 per cent, in 8 days; and to about 70 per cent, in 15 days. The red cell cholinesterase activities of several of these men were determined immediately after exposure and were found to be only slightly decreased below pre-exposure values. These results show, therefore, a correlation with the *in vitro* sensitivities of human serum and red cell activities.

Rate of Restoration of Rabbit Plasma, Red Cell, and Brain Cholinesterase Activities after Poisoning with DFP. The slow rate of regeneration of serum cholinesterase activities, demonstrated above in man and monkey, raised the question as to the rate of regeneration of brain cholinesterase activity. The average brain cholinesterase activity was first determined in a series of normal rabbits. Each of a group

of about 50 rabbits was injected with 0.3 mg. DFP per kg., and blood samples were taken, before injection, for determination of normal plasma and red cell cholinesterase. At this dose, the rabbits developed tremors within about 15 minutes after injection and continued to have these tremors throughout the day. About 10 per cent of the animals died. The surviving animals were free of symptoms the day following injection. At suitable intervals after injection, blood was taken for determination of plasma and red cell cholinesterase activities. At these or other times after injection, 1 to 6 rabbits were selected for sacrifice. Brains were removed, within 5 minutes after death, and the brain cholinesterase activities determined. The plasma cholinesterase activity returned to normal values in about 5 days. The red cell cholinesterase activity returned to normal somewhat more slowly. It was about 50 per cent of normal in about 5 days and attained the normal, pre-injection level in 10 days. The rate of recovery of brain cholinesterase activity was exceedingly slow. Ten days after injection, the brain cholinesterase activity was about 50 to 60 per cent of normal. Twenty to thirty days after injection, it was about 60 to 70 per cent of normal. Fifty days after injection, the brain cholinesterase activity had returned to 90 per cent of normal.

Attempts at Reversal of DFP Inhibition of Cholinesterase. It has been shown that the inhibitions of phosphatase by amino acids,[38] of pepsin by proteolytic digestion products,[39] and of cholinesterase by physostigmine[22, 33] may be reversed by subjecting the enzyme-inhibitor mixture to dialysis or dilution. In the present study, the serum and brain extracts of rabbits injected with 0.3 mg./kg. DFP were dialyzed against several changes of saline, for about 24 hours. Rabbit plasma was also treated *in vitro* with DFP and then dialyzed, for 24 hours. In neither type of experiment was there any increase in activity of the inactivated cholinesterase. The *in vitro* results are in agreement with those of Mackworth.[37] Dilution of mixtures of cholinesterase and fluorophosphate failed to show any relative increase in enzyme activity. This was in contrast to the results obtained on dilution of physostigmine-cholinesterase mixtures.

In view of the difference among the cholinesterases of different tissues to inhibition by DFP, it would be unjustified to draw any conclusions from our data concerning the sensitivity to inhibition of cholinesterases, at autonomic effector organs, ganglia, or myoneural junctions. According to the concept of chemical transmission of nervous impulses, the extent of cholinesterase inhibition, at these sites, should be corre-

lated with the appearance of various cholinergic symptoms. In general, in the monkey or rabbit, such cholinergic symptoms as muscular tremors, salivation, and diarrhea were associated with low red cell and brain cholinesterase activity, and death was associated with zero brain cholinesterase activity. However, this association is to be regarded as fortuitous. Conversely, it should be emphasized that depression of serum cholinesterase activity does not necessarily indicate the appearance of cholinergic symptoms. In monkey and man, for example, the serum cholinesterase activity may be reduced to extremely low levels, without the manifestations of such symptoms.

The persistence of low serum, red cell, and brain cholinesterase activity in the rabbit, for periods of 5, 10, and 60 days, respectively, and of low serum cholinesterase activities in the monkey and man, for periods of at least one to two weeks, offers evidence in support of the irreversibility of inactivation *in vivo*. Hall and Ettinger[40] found that, after injections of physostigmine in the dog, the serum cholinesterase activity dropped to 10 to 25 per cent of normal in about a half hour and returned to normal in two hours. This prompt restoration of normal activity may well be expected in the case of a readily reversible inhibitor-enzyme complex. On the other hand, the long periods of time necessary for the restoration of normal cholinesterase activity, after exposure to DFP vapor or injections with DFP solutions, in the instances mentioned, are of the same order of magnitude as those necessary for the regeneration of protein,[41] and would seem to indicate a synthesis of enzyme protein.

DFP has already been proved to be of considerable value as a tool in investigative work. Its anticholinesterase action has made it a candidate for clinical trials in *Myasthenia gravis* and glaucoma.[42–44] The marked degree to which it may inhibit cholinesterase activity *in vivo*, and the character of this inhibition, have, as we have seen, permitted studies of the rate of regeneration of various tissue cholinesterases. They open the way to further studies on the way in which diet, drugs, or the existence of various pathological lesions influence the regeneration of cholinesterase. We have also seen that DFP has permitted more incisive studies into the role of cholinesterase in nerve transmission.[45, 46]

SPECIFICITY

Specificity of Cholinesterase Action. The question of the specificity of the cholinesterase activities of various tissue extracts has claimed considerable attention. Although certain general distinctions between

the pattern of certain of the cholinesterases may be made, it will be shown that there are a number of exceptions to any strict classification.

Easson and Stedman[17] proposed the following criteria for the specificity of the cholinesterase action of different sera: (a) relative action of the serum on choline and non-choline esters; (b) inhibition of the ester-splitting action by prostigmine; (c) hydrolysis of mixed substrates. Thus, human serum was considered to contain a specific cholinesterase, because its action on non-choline esters was about 1/80th of that towards butyrylcholine, and because both actions were inhibited to the same extent by the same concentration (10^{-7} M) of prostigmine. On the other hand, guinea pig serum was considered to contain a specific cholinesterase and a non-choline, ester-splitting esterase; because the rate of actions on butyrylcholine and methylbutyrate were of about the same magnitude; because 10^{-7} M prostigmine inhibited, markedly, only the action on butyrylcholine; and because the actions on a mixture of butyrylcholine and methylbutyrate were equal to the sum of the action on each. We have found that judgment as to the relative action of a serum or tissue extract on acetylcholine and non-choline esters may depend considerably on the particular esters employed. For example, if the actions of human and rabbit brain extracts on acetylcholine are compared with that on monoacetin, it is found that there is relatively little hydrolysis of the latter. It might be concluded that these extracts contain, chiefly, specific cholinesterase. On the other hand, if the action is compared with that on triacetin, it is found that there is considerable hydrolysis of this ester, and it might just as readily be concluded that the content of non-specific esterase is very high.

Mendel and Rudney[18] stated that there were two different cholinesterases in the body: one of which acted exclusively on choline esters; and the other, a non-specific enzyme, which split both choline and non-choline esters. These they termed "true" and "pseudo"-cholinesterases, respectively. In addition to some of the criteria for specificity employed by Easson and Stedman,[17] Mendel and Rudney[18] pointed out that the non-specific or "pseudo"-cholinesterases exhibited maximal activity at high concentrations of the substrate, acetylcholine, whereas the "true" cholinesterase showed optimal activity at low substrate concentrations. According to these authors, inhibition of both choline and non-choline ester hydrolysis constitutes a criterion for distinguishing the cholinesterase as "pseudo." We have found that the anti-cholinesterase compound, DFP, markedly inhibits the mouse brain hydrolyses of both acetylcholine and triacetin. According to the criteria of Mendel and Rudney, this finding should classify mouse brain

cholinesterase as a "pseudo"-cholinesterase. Yet, these authors had classified it as a "true" cholinesterase, since its hydrolysis of acetylcholine was inhibited by physostigmine, whereas the hydrolysis of non-choline esters was unaffected by this compound.

According to Mendel's and Rudney's criteria, the red cells of several species (human, horse, sheep, ox) contained two enzymes: a specific or "true" cholinesterase and a non-choline ester-splitting enzyme. Mendel and Rudney classified red cell cholinesterase as "true" cholinesterase, because a purified preparation hydrolyzed acetylcholine, but failed to hydrolyze non-choline esters, as exemplified by methylbutyrate or tributyrin. We have found that such a purified preparation splits triacetin, a non-choline ester, which Mendel and Rudney did not test.

Nachmansohn and Rothenberg[14] have inclined towards the view that specificity is relative, and that tissue extracts containing the specific cholinesterase split acetylcholine at a higher rate than other esters. According to these investigators, the esterase in all nerve tissue is either exclusively or predominantly cholinesterase. Results which we have obtained confirm those of Nachmansohn and Rothenberg, except for the ester (triacetin), which was not tested by these investigators. This ester was hydrolyzed more rapidly than acetylcholine.

The results on several aspects of the action and inhibition of various cholinesterases have been summarized in TABLE 10. Although the data are not complete, it may be seen that, except in two respects, the cholinesterases may be divided into two general groups. In the first group, the enzymes follow the Michaelis-Menten formulation for the relationship between reaction velocity and substrate concentrations. There is no inhibition by caffeine; a marked inhibition by percaine; no activation by Na^+ and K^+; and failure to hydrolyze acetyl-B-methyl choline. In the second group, inhibition occurs at higher substrate concentrations; there are also: inhibition by caffeine; slight inhibition by percaine; activation by Na^+ and K^+; and the ability to hydrolyze acetyl-B-methyl choline. However, as already pointed out, DFP inhibits the non-choline ester hydrolysis of the enzymes, in both of these groups. Although DFP seems to inhibit the cholinesterases of the first group more markedly, there is considerable variation in sensitivity to inhibition. Moreover, there is no sharp distinction between the enzymes of these two groups, with respect to the ratio of velocities at which they hydrolyze triacetin and acetylcholine. This latter finding may be explained by assuming that there are varying amounts of non-acetylcholine hydrolyzing esterases in these preparations. Further work is

indicated to explore the actions of various cholinesterases, with respect to the criteria indicated in Table 10.

IN VIVO CHANGES OF CHOLINESTERASE

Alterations of Cholinesterase Activity in Disease. Considerable clinical and pharmacological investigation into the *in vivo* changes of cholinesterase activity has been conducted, with a view to determining the physiological significance of cholinesterase. The pathological condition which has attracted most attention, in this respect, is *Myasthenia gravis*. In this condition, which is characterized by muscle weakness and inclination to fatigue, it has been postulated that there is a deficiency of acetylcholine at the neuromuscular junction. Such a postulated deficiency may, of course, be brought about by a failure to synthesize acetylcholine, or by an excessive amount, or excessive activity, of cholinesterase at the neuromuscular junctions. That the latter mechanism is operative has been assumed, because of the finding that prostigmine, an inhibitor of cholinesterase activity *in vitro* and of serum cholinesterase activity *in vivo*, results in clinical improvement.[48] There is, however, no conclusive evidence of increased cholinesterase concentration or activity at the myoneural junctions. A number of investigators[49–55] have failed to find increased cholinesterase activity in the serum in *Myasthenia gravis*, and, although the cholinesterase activity of muscle may perhaps not be regarded as too specific, there has been a similar inability to find increase of cholinesterase activity in muscle.[56] Other explanations of the physiological fault in *Myasthenia gravis* and of the action of various drugs have been submitted by Gammon, Harvey, and Masland.[57]

There are, however, several conditions in which definite changes in serum cholinesterase activity have been reported. There is fairly general agreement that debilitating diseases, such as tuberculosis, cancer, and liver disease, are characterized by low serum cholinesterase activities.[58–64] For example, Faber[64] found ranges of 65 to 150 units in normal men and 57 to 184 in normal women. In acute hepatitis, values ranging from 41 to 51 units were obtained; in liver cirrhosis, activities ranging from 77 to 92; in cancer, from 33 to 99; and in uremia, from 32 to 56 units. High serum cholinesterase activities have been reported in hyperthyroidism.[64, 65]

The author does not know, however, of any evidence to indicate that the low serum cholinesterase values found in debilitating diseases are of any special significance, so far as transmission of nervous impulses is concerned. Indeed, these low cholinesterase activities appear to be

TABLE 10

PATTERNS OF ACTION OF VARIOUS CHOLINESTERASES

Enzyme source	Substrate-velocity relation	Caffeine inhibition	Percaine inhibition	DFP inhibition		Na, K activation	Acetyl B-methyl-choline hydroly.	Velocities ratio of TriAc ACh
				Neg. log. conc. for 50% activ.	Hydrolysis non-choline esters			
Serum: rabbit	II			4.1		+	+	6.20
horse	I	0		8.3		0		
monkey				7.7			0	0.07
human	I		+++	7.7	+		0	0.01
Red cells: human	II	++	+	5.4		+	+	0.50
Brain: human	II		±	6.0	+		+	2.20
ox		++						
mouse	II				+	+	+	0.33

merely a reflection of the state of the serum proteins. Faber[66] has found that there is a direct proportionality between serum cholinesterase activities and the concentration of serum albumin, but not between these activities and the concentration of total serum protein or that of the serum globulin. This, of course, does not imply that cholinesterase is an albumin. Faber[67] has noted that, in proteinurias, the serum cholinesterase activities remain high, relative to the concentration of the serum albumin. There is an indication, in these observations, that the formation of serum cholinesterase parallels that of serum albumin.

We have already discussed the effects of DFP on the *in vivo* activity of serum, red cell, and brain cholinesterases. There are, in the literature, similar studies on other drugs. Perhaps one of the most detailed is that of Schutz,[68, 69, 70] concerning the effect of barbiturates. This investigator noted that the prolonged administration of these drugs, in man and animals, resulted in marked decreases of serum cholinesterase activity, although these drugs in concentrations of about 0.01 M do not inhibit the *in vitro* serum cholinesterase activity. He also observed that such prolonged administration in animals resulted in *in vivo* decreases of brain, spinal cord, and muscle cholinesterase activity. His explanation was that the barbiturates decreased the activity of the cholinergic system and, hence, the demand for cholinesterase. Such an explanation is, of course, teleological and demands direct proof of decreased synthesis of the cholinesterases involved.

CONCLUSION

As a conclusion to this review, it may be of value to emphasize certain general points. The literature, as well as our own data, indicates that we cannot speak of "one" cholinesterase, identical in its properties, no matter where it may be found. Within one species, the enzyme differs, in certain respects, from tissue to tissue, and the enzyme of a given tissue may differ from species to species. Perhaps, then, it would be more proper to speak of a "family" of cholinesterases, the members of which resemble each other in some attributes and differ in others. Classification into certain groups may now be possible, but even within such groups, differences in properties may occur. Considerably more experimental work with various criteria of enzyme action is necessary, in order to achieve a more satisfactory classification. The writer believes that many of the controversies on the properties of cholinesterase which have occurred in the literature will be resolved, if the foregoing considerations are kept in mind.

It would also appear that these considerations will be of aid in elucidating various physiological and physiopathological problems. It has already been pointed out that the cholinesterase activity of the serum is not necessarily an indicator of the cholinesterase activity of the brain. Attempts to define the role of cholinesterase activity in the transmission of nerve impulses must concern themselves with a study of the properties of cholinesterase, at various loci of the nervous system, and with a correlation of the *in vivo* alterations of the activity of these cholinesterases, in response to administration of drugs or to other factors which have an effect on nerve activity. Finally, it would seem that the compound, DFP, because of its capacity for decreasing markedly, and apparently with a considerable degree of irreversibility, the content of cholinesterase in various tissues, may serve as a most useful tool for elucidating the role of cholinesterase and the factors which influence its synthesis and degradation.

BIBLIOGRAPHY

1. **Dale, H. H.**
 1914. J. Pharm. Exp. Therap. **6**: 147.
2. **Loewi, O., & E. Navratil**
 1926. Pflug. Arch. ges. Physiol. **214**: 678, 689.
3. **Michaelis, L., & M. L. Menten**
 1913. Biochem. L. **49**: 333.
4. **Lineweaver, H., & D. Burk**
 1934. J. Am. Chem. Soc. **56**: 658.
5. **Haldane, J. B. S.**
 1930. Enzymes. Longmans, Green & Co. London, New York and Toronto 85
6. **Goldstein, A.**
 1944. J. Gen. Physiol. **27**: 529.
7. **Eadie, G. S.**
 1942. J. Biol. Chem. **146**: 85.
 1941. J. Biol. Chem. **138**: 597.
8. **Wright, C. I., & J. C. Sabine**
 1943. J. Pharm. Exp. Therap. **78**: 375.
9. **Glick, D.**
 1937. Biochem. J. **31**: 521.
10. **Glick, D.**
 1938. J. Gen. Physiol. **21**: 431.
11. **Alles, G. A., & R. C. Hawes**
 1940. J. Biol. Chem. **133**: 375.
12. **Zeller, E. A., & A. Bissegger**
 1943. Helvet. Chim. Acta **23**: 1619.
13. **Mendel, B., & H. Rudney**
 1943. Biochem. J. **37**: 59.
14. **Nachmansohn, D., & M. A. Rothenberg**
 1945. J. Biol. Chem. **158**: 653.
15. **Mendel, B., & H. Rudney**
 1945. Science **100**: 499.

16. **Brown, G. L., & W. Feldberg**
1936. J. Physiol. **88**: 265.
17. **Bernheim, F., & M. L. C. Bernheim**
1936. J. Pharm. Exp. Therap. **57**: 427.
18. **Glick, D.**
1937. J. Gen. Physiol. **21**: 289.
19. **Nachmansohn, D.**
1940. Nature **145**: 513.
20. **Massart, L., & R. P. DuFait**
1939. Bull. Soc. Chim. **21**: 1039.
21. **Glick, D.**
1941. Nature **148**: 662.
22. **Straus, O. H., & A. Goldstein**
1943. J. Gen. Physiol. **26**: 559.
23. **Easson, L. H., & E. Stedman**
1936. Proc. Roy. Soc. London B **127**: 142.
24. **Collier, H. B., & D. E. Allen**
1942. Canad. J. Res. **20B**: 189.
25. **Ellis, S., F. L. Plachte, & O. H. Straus**
1943. J. Pharm. Exp. Therap. **79**: 295.
26. **Mazur, A., & O. Bodansky**
1946. J. Biol. Chem. **163**: 261.
27. **Massart, L., & R. P. DuFait**
1940. Enzymol. **9**: 364.
28. **Glick, D.**
1939. J. Pharm. Exp. Therap. **65**: 339.
29. **Zeller, E. A.**
1942. Helvet. Chim. Acta **25**: 216, 1099.
30. **Nachmansohn, D., & H. Schneeman**
1945. J. Biol. Chem. **159**: 239.
31. **Torda, C., & H. G. Wolff**
1944. Proc. Soc. Exp. Biol. & Med. **57**: 236.
32. **Heim, F., & W. Rhode**
1943. Arch. Exp. Path. Pharm. **202**: 216.
33. **Matthes, J.**
1930. J. Physiol. **70**: 338.
34. **Lange, W., & G. v. Krueger**
1932. Ber. **65**: 1598.
35. **McCombie, H., E. D. Adrian, B. A. Kilby, & M. Kilby**
1941. Personal communication.
36. **Adrian, E. D., B. A. Kilby, & M. Kilby**
1942. Personal communication.
37. **Mackworth, J. F.**
1942. Personal communication.
38. **Bodansky, O.**
1933. J. Biol. Chem. **101**: 641.
1936. J. Biol. Chem. **115**: 101.
39. **Northrop, J.**
1921. J. Gen. Physiol. **2**: 487.
40. **Hall, G. E., & G. H. Ettinger**
1937. J. Pharm. Exp. Therap. **59**: 29.
41. **Kerr, W. J., & G. H. Whipple**
1919. J. Physiol. **47**: 356.
42. **Leopold, I. H., & J. H. Comroe**
1946. Fed. Proc. **5(II)**: 190.

43. **Comroe, J. H., J. Todd, G. Gammon, G. B. Koelle, & A. Gilman**
1946. Fed. Proc. **5**(II): 172.
44. **Harvey, A. M., B. F. Jones, S. Talbot, & D. Grob**
1946. Fed. Proc. **5**(II): 182.
45. **Rothenberg, M. A., & D. Nachmansohn**
1946. Fed. Proc. **5**(II): 199.
46. **Crescitelli, F., G. B. Koelle, & A. Gilman**
1946. Fed. Proc. **5**(II): 172.
47. **Easson, L. H., & E. Stedman**
1937. Biochem. J. **31**: 1723.
48. **Walker, M. B.**
1935. Proc. Roy. Soc. Med. **28**: 759.
49. **McGeorge, M.**
1937. Lancet **1**: 69.
50. **Milhorat, A. T.**
1938. J. Clin. Invest. **17**: 649.
51. **Freudenberg, R., & F. K. Redlich**
1938. Arch. Exp. Path. Pharm. **188**: 645.
52. **Poncher, H. B., & H. W. Wade**
1939. Arch. Neurol. Psychiat. **41**: 1127.
53. **Stedman, E., & W. R. Russell**
1937. Biochem. J. **31**: 1987.
54. **Odom, G., C. K. Russell, & D. McEachern**
1943. Brain **66**: 1.
55. **Stoner, H. B., & A. Wilson**
1943. J. Physiol. **102**: 1.
56. **Jones, W. C., & W. C. Stadie**
1939. Quart. J. Exp. Physiol. **29**: 63.
57. **Gammon, G. D., A. M. Harvey, & R. L. Masland**
1941. Biol. Symposia **3**: 291.
58. **Glick, D.**
1941. Biol. Symposia **5**: 213.
59. **Antopol, W., A. Schifrin, & L. Tuchman**
1938. Proc. Soc. Exp. Biol. & Med. **38**: 363.
60. **McArdle, B.**
1940. Quart. J. Med. **9**: 107.
61. **Vahlquist, B.**
1935. Skand. Arch. Physiol. **72**: 133.
62. **Scoz, G., & C. Cattaneo**
1937. Enzymol. **4**: 157.
63. **Vincent, D., & J. de Prat**
1942. C. R. Soc. de Biol. **136**: 821.
64. **Faber, M.**
1943. Acta Med. Skand. **114**: 59.
65. **Antopol, W., L. Tuchman, & A. Schifrin**
937. Proc. Soc. Exp. Biol. & Med. **36**: 46.
66. **Faber, M.**
1943. Acta Med. Skand. **114**: 72.
67. **Faber, M.**
1943. Acta Med. Skand. **115**: 475.
68. **Schutz, F.**
1943. Quart. J. Exp. Med. **33**: 35.
69. **Schutz, F.**
1941. Nature **148**: 725.
70. **Schutz, F.**
1942. J. Physiol. **102**: 259, 269.

THE EFFECTS OF DRUGS ON NERVE ACTIVITY

By Alfred Gilman*

Pharmacology Section, Medical Division, Chemical Warfare Service, Edgewood Arsenal, Maryland

I cannot help but feel somewhat apologetic for having consented to contribute to this symposium. Indeed, as the date for this conference approached, it became more and more apparent that I had become involved in an extremely paradoxical situation. On first thought, it would seem reasonable to request a pharmacologist to discuss the subject of the effects of drugs on nerve activity. However, a moment's reflection, which unfortunately, from my point of view, was too long delayed, results in the irrefutable conclusion that it is the neurophysiologist who should be addressing the pharmacologists on this subject.

Although the pharmacologist has, as his ultimate objective, the elucidation of the fundamental mechanism of action of drugs on cells, he is continually frustrated by the limitations of his own technics. The very nature of the subject of pharmacology, which borders on so many medical disciplines, almost precludes the possibility of the investigator in this field engaging in the basic research which is essential for the reaching of his objective, except, possibly, in a chosen, limited field. It is from the neurophysiologist, therefore, that the answer to many of the basic problems of the actions of drugs on the nervous system can be expected.

If one wishes to indulge in oversimplification, the entire subject of the effects of drugs on nerve activity can be summarized in a few moments or even in a single sentence. There is no phase of nerve activity which cannot be profoundly affected by drugs. Effector cells can be completely released from nervous control or, conversely, the effects of nerve impulses can be faithfully mimicked; conduction in nerve fibers can be completely blocked; synaptic transmission can be interrupted or enhanced; cord transection can be simulated; selected centers in the brain can be stimulated or depressed. It only remains to name the drugs associated with these actions.

Any further amplification would result in a textbook discussion, in which, in a more or less orderly fashion, the actions of drugs could

* Major, Sn-C, A. U. S.

be classified in such general terms as local and general anesthetics, analgesics, central stimulants, blocking agents, etc. For the purpose of understanding the therapeutic application of drugs modifying nervous activity, such information may be adequate. However, from the standpoint of the contributions that drugs can make toward the solution of basic problems in neurophysiology, our understanding of their mechanism of action is as yet inadequate.

Drugs have proved to be invaluable tools in many fields of biological and medical research, a statement which the neurophysiologist would be the last to deny. Their value is readily appreciated, when one considers one of the few basic and general statements which can be made concerning the fundamental mechanism of action of drugs: namely, that drugs cannot impart new functions to cells or tissues, but can only modify, *i.e.*, stimulate, depress, or block, functions which are the fundamental properties of that cell or tissue. Thus, when a drug produces general or local anesthesia, has a convulsant action, blocks synaptic transmission, stimulates chemoreceptors, or affects nervous activity in any of a variety of ways, no matter how extreme the response, the assurance is justified, until an exception to the general rule can be proved, that the drug in question has modified a normal cellular function.

Examples of how chemical agents, foreign to the body, have contributed to physiological concepts are numerous. Indeed, through these the subject matter of neurophysiology has been enriched. Surely, it cannot be mere coincidence that so many basic contributions to the concept of the chemical transmission of the nerve impulse had their origin in pharmacological laboratories? Rather, the knowledge that chemical agents could mimic, in end-organs, the effect of nerve stimulation served as the basic stimulus for the search for evidence of chemical mediation. Is it not possible that, in some drug, still inadequately explored by the neurophysiologist, lies the answer to a basic neurophysiological problem? As a possible example, let us consider the local anesthetics. The local anesthetic action of cocaine was first demonstrated in 1884. This type of action has been shown to be exhibited by a variety of chemical structures, but the tertiary amino esters of benzoic acid and para-amino benzoic acid, as a group, are capable of blocking transmission. These compounds show no respect for any classification of nerves, but block cholinergic and adrenergic, sensory and motor fibers in an indiscriminate manner, which points to a basic action on a fundamental mechanism of transmission, shared by all nerves. Is it not reasonable to suppose that, by inquiring more

deeply into the action of drugs such as the local anesthetics on a cellular level, information may be gained on the mechanism of the propagation of the nerve impulse?

The full realization of the contributions that drugs can make to the elucidation of fundamental physiological mechanisms, can only result from the cooperative research efforts of the pharmacologist and those investigators who are focusing their interests more intensively on a specific field. This is readily appreciated, in so far as the full exploitation of the therapeutic applications of a drug is concerned. The pharmacologist, who, in the course of an investigation on central depressants, finds a compound with new and significant anti-convulsive properties, will invariably refer the compound to a clinical neurologist, if he deems it worthy of consideration as an anti-epileptic. During the course of chemical warfare research, compounds were studied which, to the pharmacologist, suggested themselves as potential therapeutic agents in the treatment of such unrelated conditions as Hodgkin's disease, mercury poisoning, *Myasthenia gravis*, and glaucoma. In every instance, the prediction of therapeutic worth which was based upon laboratory analyses proved correct, but it was only through the efforts of clinicians, highly specialized in their particular field, that the full appreciation of the therapeutic value of these particular agents was realized.

Drugs are constantly following the path from laboratory to clinic, and many of the outstanding medical accomplishments of the past decade are the result of this cooperative effort. However, drugs can only make their full contribution to the science of medicine when they also follow another, more fundamental, and possibly more important path: namely, from the laboratory of the pharmacologist to the laboratories of investigators working on those physiological problems which attempt to define biological processes in their most fundamental terms. That drugs are not being fully exploited, in this respect, is a regrettable fact. Even in the field of neurophysiology, where drugs have proved such valuable research tools, a full realization of their potential contributions has not been reached. The tendency, rather, has been to accept drugs with known actions and to employ them for these actions, rather than to investigate unknown mechanisms of drug action as probes into physiological processes. This occurs despite the fact that acetylcholine and physostigmine, two drugs which are better understood than any other compounds affecting nerve action, have paved the way toward revolutionary concepts in an understanding of synaptic transmission.

New agents affecting the nervous system are constantly being de-

scribed. They are conveniently labeled as convulsants, depressants, etc., and sporadic attempts are made to localize their action, following which, investigation is considered to be complete. I wish to take this opportunity to present three new compounds which have come to the attention of those who, during the past few years, have been working in the field of chemical warfare. These are highly toxic agents which must, by definition, exert profound effects on basic cellular mechanisms. As a result of the cooperative research program, associated with the war effort, these compounds have received more fundamental study, in the course of a few years, by groups with more divergent interest, than is usually the case with better-known and more widely-employed drugs. These three compounds have been selected for discussion from a large group, because of the profound actions which they exert on the nervous system.

The first of these compounds is the sodium salt of fluoroacetic acid. This agent has had an interesting history. It represents, in its practical applications, one of the many fruitful by-products of chemical warfare research. Sodium fluoroacetate was screened by the Fish and Wildlife Service and has proved to be by far the most outstanding of all rodenticides. Now known as Compound 1080, it promises to be an important agent in the control of the spread of disease and the economic wastage caused by rodents. Yet the prediction is not unjustified that the compound may prove to be of even more significant value as a laboratory research tool, assuming equal importance, in this respect, with the iodoacetates.

Although the fluoroacetate ion resembles iodoacetate in structure, it shares none of its chemical or physiological properties. However, both halogenated acetates exert profound effects on cellular metabolism. Highly provocative are the observations of Barron and co-workers,[1] that the oxidation of acetate by yeast, rat kidney suspensions, and heart slices is inhibited by fluoroacetate. When pyruvate is used as the oxidizable substrate, there is an accumulation of acetate, after the addition of fluoroacetate, and the synthesis of carbohydrate from pyruvate is completely inhibited. Barron has advanced the reasonable hypothesis that fluoroacetate, because of its close structural relationship to acetate, blocks, by competitive inhibition, enzyme systems concerned with the utilization of acetate.

When one considers the basic importance of acetate metabolism, it is of extreme interest to inquire into the pharmacological actions of a compound which, possibly, interferes with the utilization of this essential metabolite. These actions have been investigated by Chenoweth

and Gilman.[2] Species vary greatly, both in their response and tolerance to fluoroacetate. In general, the fluoroacetate ion possesses two main pharmacological actions. In some animals, it affects primarily the heart. Energy metabolism and conduction are so disturbed that *pulsus alternans*, A-V block, frequent ectopic ventricular beats, and eventually, ventricular fibrillation result. Most herbivorous animals, as well as those species of primates that have been studied, are susceptible to the cardiac actions of the fluoroacetates. The actions of the fluoroacetates on the nervous system are even more striking. Following a latent period of approximately one hour, animals become progressively more excited, and eventually exhibit severe epileptiform convulsions which continue, uninterrupted, until death. Dogs and cats respond in this manner. Most of those species which exhibit the cardiac actions show no evidence of central stimulation. However, primates, although succumbing to the cardiac action, may show mild epileptiform convulsions. In regard to susceptibility, the lethal convulsive dose in the dog is approximately 0.1 mg./Kg. The lethal dose in primates is approximately 100 times as great.

When one considers the descriptive pharmacology of fluoroacetate in the light of its possible fundamental mechanism of action, certain questions come immediately to mind. Are differences in species response due to different metabolic patterns in their nervous tissue? Observations of Chenoweth and co-workers support the view that the primate myocardium is uniquely dependent upon the utilization of acetate, for adequate function. Similar investigations may reveal that the same is true of the nervous system in the case of the dog and cat. It has been shown by Tepperman and Mazur[3] that, in the presence of fluoroacetate, acetylation is greatly enhanced, presumably because of the high concentration of available acetate. Could this finding possibly be concerned with the convulsive action of the fluoroacetate ion? These are but a few of the problems, pertinent to the nervous system, that have been raised by the preliminary investigations of this drug. By the proper utilization of fluoroacetate as a research tool, it may be possible to relate specific disturbances in the metabolism of nervous tissue to functional abnormalities. In this respect, it is of interest to note that the electroencephalogram obtained during a fluoroacetate-induced convulsion is almost identical to that of a *petit mal* epileptic seizure.

The second compound to be discussed is diisopropyl-fluorophosphate. This compound represents a new type of anticholinesterase. Not only does it depart, in its chemical configuration, from previously studied

anticholinesterase agents, but it also differs in its fundamental mechanism of action, in that the inhibition of cholinesterase is irreversible. What is more, diisopropyl-fluorophosphate is highly lipoid-soluble and rapidly gains access to nervous tissue.

In diisopropyl-fluorophosphate, the neurophysiologist has at his command a research tool in which the fundamental mechanism of action is known. Thus, if the major premise of my introductory remarks is to hold true, the application of this compound to problems of neurophysiology should help to prove or disprove fundamental concepts of nerve function.

The advantages of an irreversible anticholinesterase, as a research tool, are at once evident. Following the action of diisopropyl-fluorophosphate, the cholinesterase activity of a tissue can be restored only by resynthesis of enzyme. Moreover, the agent can be administered, and the response of a tissue studied. Following this, the tissue can be removed and ground; the homogenate appropriately diluted; and the absolute cholinesterase activity determined; an approach which cannot be employed with a compound such as physostigmine, due to the reversible nature of its inhibition. Thus, for the first time, a highly quantitative approach to the problems of the role of cholinesterase and acetylcholine in the transmission of the nerve impulse is available.

Diisopropyl-fluorophosphate has already followed the path from the laboratory to the clinic. Comroe and associates,[4] at the University of Pennsylvania, and Harvey and co-workers,[5] at Johns Hopkins, have employed this agent in the treatment of *Myasthenia gravis*. The therapeutic efficacy of this type of compound, as well as its limitations, has already been demonstrated. Of even greater interest, will be the more fundamental data, from these studies, which may shed light on the defect in transmission associated with this myopathy.

Basic laboratory studies, employing diisopropyl-fluorophosphate as a research tool, have also begun. I should like to report, in some detail, the experiments of Crescitelli and co-workers,[7] designed to elucidate the possible role of acetylcholine in the conduction of the nerve impulse along the nerve fiber. The background literature to this problem has recently been summarized by Loewi[8] and by Feldberg,[9] and need not be repeated here. Mention should be made, however, of the studies of Cowan,[9] Lorente de Nó,[10] Hertz,[11] and Cantoni and Loewi,[12] in which either physostigmine or acetylcholine failed to exert a significant effect on transmission in the nerve fiber. However, the availability of an irreversible inhibitor of cholinesterase, which afforded an opportunity to correlate nerve function with quantitative data on cholin-

esterase concentration, prompted a repetition of this type of study. Moreover, advantage was taken of the opportunity to compare the effects of a reversible (physostigmine) and an irreversible anticholinesterase. The nerve action potential was employed as an index of effect on transmission. It was argued that, whereas both types of anticholinesterase agent should affect the nerve action potential in the same manner, assuming acetylcholine to play a major role in transmission, the action of physostigmine should prove reversible, that of diisopropyl-fluorophosphate, irreversible. The possibility that physostigmine might not gain access to those structures accessible to diisopropyl-fluorophosphate was avoided, as far as possible, by employing the alkaloid base, as well as the salicylate salt.

Two types of experiments were performed. In the first, isolated nerves of bull frogs and of cats were mounted in a moist chamber, placed in a constant temperature bath of appropriate temperature. A portion of the nerve was looped into a small chamber containing Ringer's solution. Following the recording of control action potentials, the effects of the various drugs were ascertained. When the isolated nerves of the cat or the bull frog were exposed to 0.01 molar physostigmine salicylate, no detectable change in the action potential was observed. However, when the solution containing the salicylate salt of physostigmine was replaced by the alkaloidal base, the action potential disappeared within a period of 10 minutes. Washing the nerve with Ringer's solution restored the action potential completely. Thus, the water-soluble salicylate salt was devoid of action, whereas the lipoid-soluble alkaloidal base blocked transmission.

Similar experiments were then performed, by exposing the nerve to 0.02 molar diisopropyl-fluorophosphate. Again, the action potential disappeared within a few minutes. It only remained to demonstrate the irreversibility of this block, in order to attribute the effect to the inactivation of cholinesterase. However, washing the nerve restored the action potential, despite the fact that the action of diisopropyl-fluorophosphate was supposedly irreversible. In view of this surprising result, experiments were performed to determine the extent of the wash necessary to restore the nerve and the speed at which the action potential returned. During the course of these studies, it was observed that it was only necessary to remove the nerve from contact with the solution containing diisopropyl-fluorophosphate, to restore the action potential completely. Thus, the conduction defect could not have been related to an inhibition of cholinesterase.

It was not possible, in the experiments on the isolated nerve, to measure accurately the extent of inhibition of cholinesterase, for the reason

that only a very small segment of nerve was exposed. It was, thus, impossible adequately to wash the nerve, so as to preclude the possibility of mechanical transfer of sufficient diisopropyl-fluorophosphate to inactivate cholinesterase during the preparation of the nerve for enzymatic studies. For this reason, a series of experiments was performed in which a large dose of diisopropyl-fluorophosphate was injected into the ventral lymph sac of frogs and allowed to reach the nerve, by way of the circulation. After a suitable interval, the nerves were dissected and their transmission characteristics studied and compared with control frogs. Following this, the cholinesterase content of the control and experimental nerves was studied. Despite the fact that the nerves of the experimental frogs were completely devoid of cholinesterase, the transmission of the nerve impulse, as determined by the characteristics of the action potential, in response to single and repetitive stimuli, was unaffected. This finding casts serious doubt on the role of acetylcholine as a de-polarizer, in the processes of conduction along the axon

Loewi, in his recent review, quotes Dale as having once remarked that it was unreasonable to suppose that nature would provide for the liberation in the ganglion of acetylcholine, the most powerful stimulant of ganglionic cells, for the sole purpose of fooling physiologists. What, then, is the function of cholinesterase in nerve fibers, which Nachmansohn and his co-workers have shown so conclusively to be concentrated at the surface, rather than in the axoplasm? The answer is not yet forthcoming. However, in a drug like fluorophosphate, it is possible, by localized injection, to reduce the concentration of cholinesterase in a chosen tissue to negligible amounts. Thus, we have a research tool which may provide the answer to these basic problems.

The third agent will be discussed only very briefly. It shares with diisopropyl-fluorophosphate the ability irreversibly to inactivate cholinesterase. It differs from diisopropyl-fluorophosphate in possessing a more outstanding action on the nervous system Certain species, in particular cats and dogs, exhibit severe convulsions, which have their onset within a few minutes after the intravenous injection of the drug and which persist until death. The fact that an anticholinesterase agent possesses such extreme convulsant action could possibly be attributed to coincidence. However, there is one finding which points to an intimate relationship between convulsions and the chemical mediation of central synaptic transmission. If the animals receive a therapeutic dose of atropine, before the administration of this anticholinesterase, no convulsions are observed, and complete protection

is afforded from what would, otherwise, be a lethal dose. Moreover, if the agent is administered and the convulsions are allowed to progress to their peak intensity, the intravenous injection of atropine stops all convulsive activity within 30 seconds, and the animal appears normal, as soon as it recovers from its exhaustion. It should be emphasized that the doses of atropine that exert this anticonvulsant action are of a small magnitude and, in themselves, exert no demonstrable central effects. No other central stimulant can be inhibited in this manner. The effect of atropine, in blocking the convulsant action of this anticholinesterase, is as dramatic and as complete as is the effect of atropine in blocking the reception of post-ganglionic cholinergic impulses by autonomic effector cells. It seems certain that, in this compound, there is a research tool which can make a significant contribution to the fundamental problems of central synaptic transmission.

During the past few years, the group of investigators at Edgewood Arsenal has been engaged in a cooperative research effort, in which toxic war gases or, in other words, highly active drugs, were the focal point of their investigations. Their efforts were coordinated with extensive programs of numerous academic groups. From the point of view of the pharmacologist, this elaborate approach to the mechanism of drug action has proved to be an illuminating experience. One cannot fail to be impressed by the fact that, as the story of each agent unfolded, its potential value toward the solution of fundamental problems in physiology and biochemistry was more and more appreciated.

I have departed from a routine discussion of the effects of drugs on nervous activity, to present to you three new agents which have resulted from this program. It may be predicted that, as research tools, they will prove invaluable. If so, then the contention made earlier will have been fulfilled: that, by tracing the actions of drugs to their cellular mechanisms, basic physiological processes will be revealed.

REFERENCES

1. **Barron, E. S. G., G. R. Bartlett, & G. Kalnitsky**
1946. Fed. Proc. 5(II): 120.
2. **Chenoweth, M. B., & A. Gilman**
1946. J. Pharm. Exp. Therap. **87**: 90.
3. **Tepperman, J., & A. Mazur**
Personal Communication.
4. **Comroe, J. H., Jr., J. Todd, G. Gammon, G. B. Koelle, & A. Gilman**
1946. Fed. Proc. 5(II): 172.
5. **Harvey, A. McG., B. F. Jones, S. Talbot, & D. Grob**
1946. Fed. Proc. 5(II): 182.
6. **Crescitelli, F. N., G. B. Koelle, & A. Gilman**
1946. J. Neurophysiol. **9**: 24.

7. **Loewi, O.**
1945. J. Mount Sinai Hosp. **12:** 851.
8. **Feldberg, W.**
1945. Physiol. Rev. **25:** 596.
9. **Cowan, S. L.**
1938. J. Physiol. **93:** 215.
10. **Lorente de No, R.**
1944. J. Cell. Comp. Physiol. **24:** 86.
11. **Hertz, H.**
1945. J. Physiol. **104:** 1.
12. **Cantoni, G. L., & O. Loewi**
1944. J. Pharm. Exp. Therap. **81:** 67.

THE RECOVERY OF DIAMETER AND IMPULSE CONDUCTION IN REGENERATING NERVE FIBERS

By Charles M. Berry and Joseph C. Hinsey

Department of Anatomy, Cornell University Medical College, New York, N. Y.

The primary purpose of these experiments was to study those properties of regenerating fibers which could be observed oscillographically and related to histological controls. Therefore, the contours of the action potentials, the conduction velocities of the impulses, and fiber diameters were followed in a series of cat nerves. The recovery of these properties might be considered to be a process of reconstitution or maturation, as opposed to the longitudinal outgrowth of the fibers, and since these properties continued to change over long periods of regeneration, measured in years, the experiments were spread over a wide range of time, from a few days to more than three years.

Having observed the effects of time (which is undoubtedly the most important factor in the reconstitution of fibers distal to a suture), we extended the experiments to include the effects of crushing the nerves compared to section and suture; the effects of delaying the suture after transection; and the effects of cross-suturing nerves containing fibers of different fiber diameters. The importance of these factors has been reemphasized in recent publications. The growth of fiber diameters in the distal stump over a one year period has been carefully plotted by Gutmann and Sanders.[1] Furthermore, they showed differences in recovery between crushed and sutured nerves. The influences of physical stresses in the "union" tissue described by Weiss[2] show the importance of the type of junction between central and distal stumps. The effects of delaying the suture after section of a peripheral nerve were studied by Holmes and Young,[3] and the effects of cross-suturing visceral and somatic nerves were reported by Simpson and Young.[4] Young and his co-workers have paid special attention to the connective tissue sheath diameters in the distal stump.

METHODS AND RESULTS

Action Potentials from Regenerating Nerves

Cathode-ray oscillographs were taken from regenerating tibial, peroneal, and saphenous nerves of 64 cats. The nerves were transected

with a sharp blade and immediately sutured with silk. Each nerve was allowed to regenerate for a determined interval, from 17 to 1363 days, and was then excised and placed on electrodes in a moist chamber at 38° C.

The regenerating fibers were able to conduct impulses after a very brief regeneration period, and at 17 days, action potentials were recorded from the distal stump of one tibial nerve within a distance of 2 cm. from the suture This potential was small, less than 10 microvolts in amplitude, and was conducted very slowly at a maximum of 0.9 meters per second. The potentials recorded from fibers which had regenerated for longer periods were of greater amplitude and were conducted at greater velocities. The increase in conduction velocity was rapid in the first few days and, by 36 days, velocities of 17 m.p.s. were recorded, as shown in PLATE 5 A. This record was taken from a monopolar electrode placed 3.5 cm. distal to the suture, and shows a maximum conduction velocity of 17 m.p.s. and a spike amplitude of 25 microvolts. The maximum conduction velocity continued to increase with the time allowed for regeneration, but at an ever-decreasing rate of recovery. Thus, at 50 days, 25 m.p.s. were attained; at 100 days, 40 m.p.s.; at 200 days, 60 m.p.s.; at 365 days, 70 m.p.s. Beyond 544 days, no further recovery of conduction velocity was found, and at the long period of 1363 days, only 80 m.p.s. were attained. The record in PLATE 5 C was taken from the distal stump of a tibial nerve 1363 days after suture, and can be compared with the record from the opposite, normal, tibial nerve of the same animal in PLATE 5 B, in order to determine the degree of recovery. In PLATE 5, B and C, the conduction distance was 8 cm., but in the record from the regenerated nerve, the distance or time between the shock artifact and the beginning of the spike is greater than in the normal record, and shows that the 80 m.p.s. represent less than 80% recovery toward the normal conduction velocity. Similar results were obtained from the peroneal and saphenous nerves.

Two other observations can be made from PLATE 5, B and C. Firstly, the amplitude of the spike is less in the regenerated nerve, and secondly, the spike in PLATE 5 C is not as complex. The lack of recovery of all the components of the spike was even more obvious in records from the saphenous nerves, where the normal potential is more complex and consists of a double or triple peaked alpha wave and distinct beta, gamma, and delta waves. Even after long periods of regeneration, the saphenous nerve did not recover these wave components and showed only an initial peak which leveled off into a long tail.

An accurate method of determining the conduction velocities of the most rapidly conducting fibers is illustrated in FIGURE 1. Several records were taken along a regenerating nerve at various conduction distances, and either the stimulating or the recording electrodes were placed at a fixed point along the nerve. Then the other electrodes were moved stepwise, to provide a greater conduction distance for each successive record. Thus, in FIGURE 1, the stimulating electrodes were placed 4.6 cm. distal to the suture, and the pair of recording electrodes was placed at a variety of points both distal and proximal to the

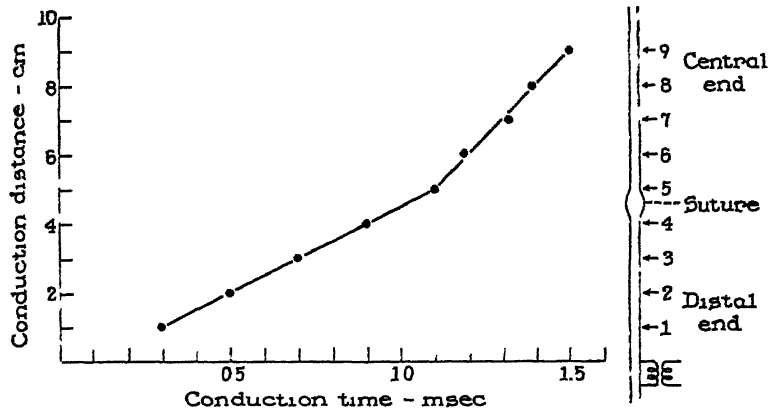


FIGURE 1 Graph of conduction distance against conduction time of the action potentials, from a saphenous nerve 58 days after suture. The diagram on the right shows that the stimulating electrodes were held stationary on the distal stump while the recording electrodes were moved. The conduction velocity jumped from 23 m.p.s. in the distal stump to 50 m.p.s. in the central stump (**Berry, Grundfest, & Hinsey.**)

suture. At each distance, a record was taken, and the conduction time was measured between the shock and the initial rise of the spike. When this time was plotted against conduction distance, as in FIGURE 1, the slope of the line indicated the maximum conduction velocity.

The change in the slope of the line, in FIGURE 1, occurred at the suture line which shows that the distal outgrowths conduct more slowly than their parent fibers in the central stump. The actual velocities from this saphenous nerve, 58 days after suture, were 23 m.p.s. distal to the suture, and these same fibers central to the suture conducted at 50 m.p.s. The continuity of the plotted line and its straight contour central to the suture demonstrates that the change in velocity was recorded from identical fibers on both sides of the suture.

Assuming that conduction velocity is related to the fiber diameter, the electrical method proves that the small fibers of the distal stump are not the result of a selective ability of only the small fibers of the

central stump to grow out into the distal stump, before the larger fibers can grow. Conversely, it demonstrates that large fibers of the central stump send out small extensions into the distal stump, which then mature and take on the action potential characteristic of smaller fibers, regardless of their origin.

A consistent finding, not illustrated in FIGURE 1, was that the conduction velocity of the distal fibers was less at greater distances from the suture.

Fiber Diameter Measurements

The regenerating tibial, peroneal, and saphenous nerves which had been excised and used for action potential experiments were fixed in osmic acid and studied microscopically. The most obvious change in the character of the regenerating nerves, as they were allowed to grow for longer and longer periods, was the gradual increase in fiber diameters. The outside diameters, including both axis cylinders and myelin sheaths, were measured with a movable, ocular micrometer. To insure random sampling, the fibers were measured in horizontal bands, with approximately 500 fibers measured in each nerve. The results were plotted along fiber distribution curves, as illustrated in FIGURE 2. These nine histograms were picked from a series of regenerating tibial nerves, to show the diameter characteristics of the fibers in the distal segments at different time intervals following suture. Since the diameters were smaller, at greater distances from the suture, the portions of tibial nerves studied in each of the histograms in FIGURE 2 were taken from similar levels, just beyond the upper branches to the gastrocnemius muscles, 3 to 5 cm. distal to the sutures.

The shift in distribution from left to right in the histograms of FIGURE 2 shows rapid diameter growth between 33, 59, and 127 days after transection and suture. At longer regeneration times of 207, 318, and 420 days, the fibers continued to mature, but more slowly. At 544, 901, and 1363 post-operative days, there was negligible increase in fiber diameter, but a complete recovery of the normal fiber size was never attained. Even at 1363 post-operative days, the large group of fibers between 9 and 20 micra had not appeared. Similar lack of complete recovery of fiber size was found in the peroneal and saphenous nerves.

Particular attention was paid to the measurement of the largest fibers in each nerve, since the maximum conduction velocity had already been determined accurately, and it was a reasonable assumption that the largest fibers were responsible for the maximum conduction velocity. The growth of the largest fibers is illustrated in FIGURE 3

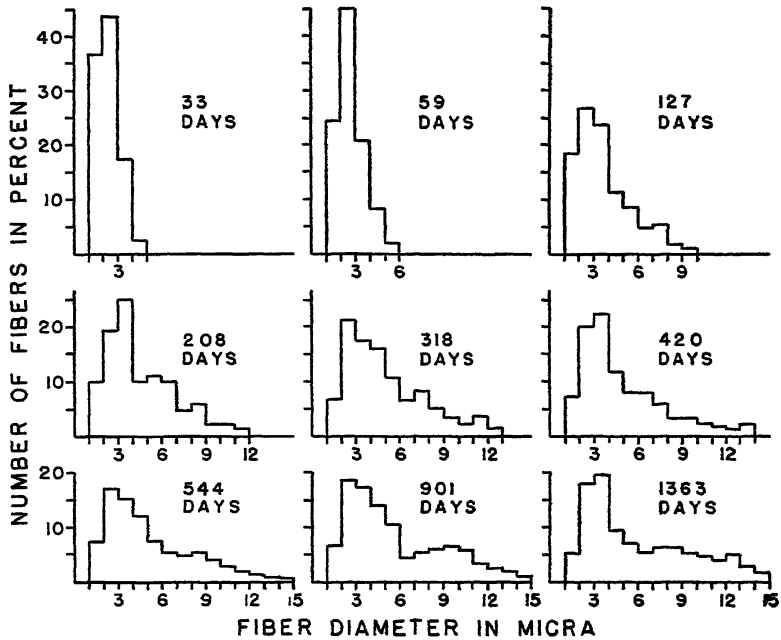


FIGURE 2 Reconstitution of fiber diameter with regeneration time. Each fiber diameter distribution graph is from a different nerve, excised on the indicated number of days after transection and suture.

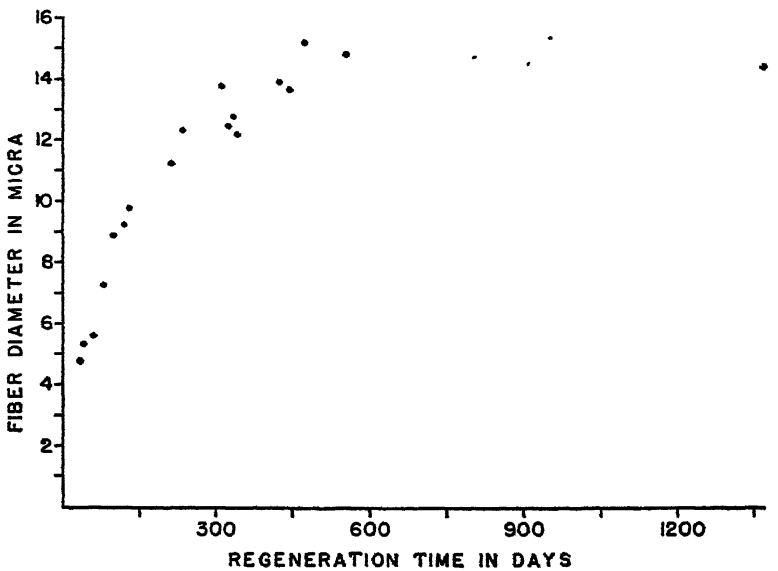


FIGURE 3. Outside diameters of the largest fibers in the distal stumps of sutured tibial nerves. The fibers do not attain the normal maximum diameters, above 20 micra.

as a graph of fiber diameter against the amount of time allowed for regeneration after suture of the tibial nerves. The leveling off of the diameter growth curve at a level between 14 and 16 micra again illustrates the incomplete "maturation" of fibers, even after long periods of regeneration.

The Relationship between the Conduction Velocity and the Fiber Diameter

The results from both the action potential and the fiber diameter studies showed a gradual recovery, which tapered off with no complete

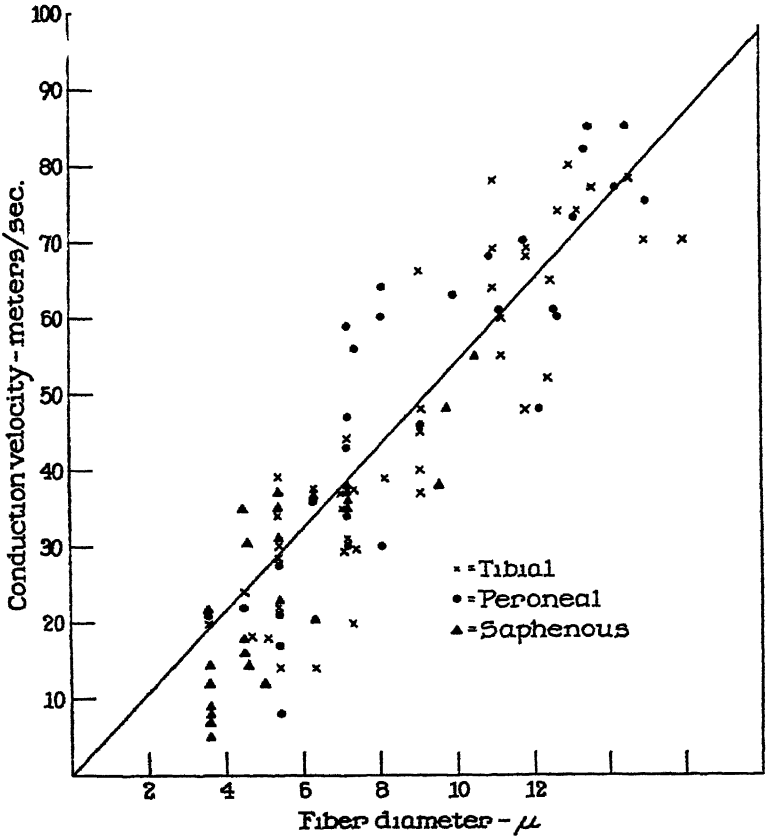


FIGURE 4. Graph showing the relationship between the maximum conduction velocities and the outside diameters of regenerating fibers of the nerves: tibial (crosses) peroneal (circles), and saphenous (triangles). (**Berry, Grundfest, & Hinsey.**[7])

return of either the velocity of impulse conduction or the size of the fibers. The actual relationship between the two functions is shown in FIGURE 4, where the maximum conduction velocity is plotted against

the maximum fiber diameter for each nerve. The linear relationship between the two functions is thus illustrated in a new way, and confirms the contention of Gasser and Grundfest,[3] that the relationship is a linear one. Less deviation from the straight line in FIGURE 4 was found at the lower end of the graph, if the inside diameter (without myelin sheath) was measured instead of outside diameter.

The Difference in Recovery between Sutured and Crushed Nerves

In a small series of 10 cats, the tibial, peroneal, and saphenous nerves were crushed with thin, flat-surfaced forceps, and the nerves were allowed to regenerate for determined intervals. The purpose of these experiments was to determine whether the recovery of the action potential and fiber diameter would occur in the same way as had been observed in the sutured nerves.

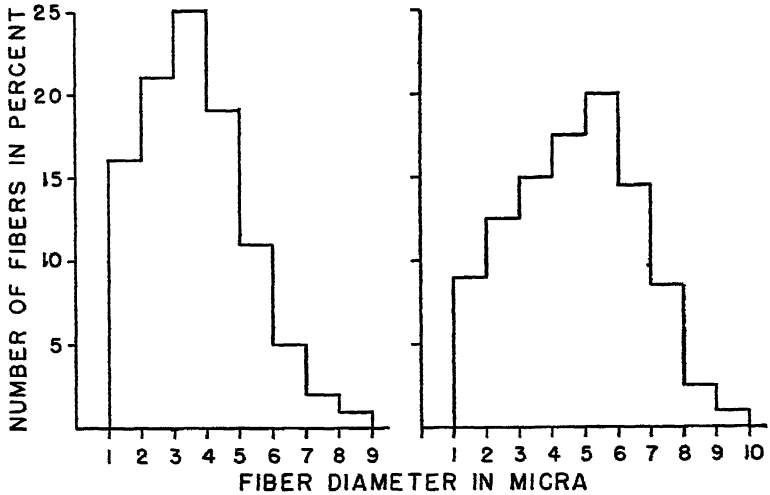


FIGURE 5 Comparison of fiber distribution according to diameter.
Histogram on the left, after suture.
Histogram on the right, after crush.
Both tibial nerves were allowed to regenerate for 94 days.

The action potential records from the crushed nerves showed greater recovery of conduction velocity and magnitude of the spikes than was found in records from comparable regions of sutured nerves, taken after the same amount of time had been allowed for regeneration. The fibers also grew in diameter more rapidly in the crushed nerves. In FIGURE 5, the histogram on the left is from the distal stump of the tibial nerve, 94 days after *suturing*, that on the right, from a tibial nerve, 94 days after *crushing*. Shift of the graph to the right, in the crushed

nerve compared to the sutured, demonstrates the more rapid recovery of fiber size in the crushed nerves. These experiments were not carried beyond 200 days, and the ultimate amount of recovery of the normal fiber distributions was not determined for crushed nerves.

The Effects of Delayed Suture on Fiber Diameter Growth

The method of studying the "maturation" of the fibers by action potential records and diameter measurements was used to determine the influence of delaying the suture after transection. An operative procedure similar to that of Holmes and Young was devised, to allow the tibial nerve to degenerate after section and remain free of fibers during a determined delay period. The adjacent peroneal nerve was then transected, and the freshly cut central stump was sutured to the old distal remnant of the tibial. This cross-suture was used to limit the study to effects of delay in the distal segment.

In the same animal, a reliable control was provided in each experiment by suturing the peroneal nerve of the opposite leg to the tibial nerve. This was done in exactly the same manner as on the delayed side, but, in this case, there was no delay between section and suture.

The procedure may be summarized as follows: The right tibial nerve was exposed, and a long segment removed from the sciatic notch to the popliteal space. To insure the absence of regeneration during the delay period, the cut nerve was exposed at 6-month intervals. Then, after a delay of 14 to 476 days, a second operation was performed, in which the right peroneal was sectioned and the central end sutured in the old distal tibial. At this same time, the left peroneal was sectioned and sutured to the distal tibial as a control. 105 to 440 days were then allowed for regeneration before the terminal experiment.

The shapes of the action potentials and the maximum conduction velocities in the distal segments, after delay periods of 14, 21, 28, 56, 84, and 180 days, were similar to those recorded from the nerves of the opposite leg which had been sutured without delay. In each of these experiments, 105 days were allowed for regeneration after suture. The diameter distribution of the fibers was the same on both sides, as illustrated in FIGURE 6, A and B. The histogram in FIGURE 6 B was from the right tibial, 3 cm. distal to the suture, which was delayed 84 days, and after which the nerve regenerated for 105 days before the fibers were measured. The control from the same animal is shown in FIGURE 6 A. Both nerves were analyzed 105 days after suture. The similarity of the two histograms was also found with 14, 21, 28, 56, and 180 days, and indicates that such delay periods had no influence on the diameter growth of fibers, under these experimental conditions.

A suture delayed for a longer period of 253 days resulted in less reconstitution of the fiber diameters in the peripheral stump than that found in the control. The histogram in FIGURE 6 D was from the distal tibial, 3 cm. distal to the suture, which was delayed 253 days.

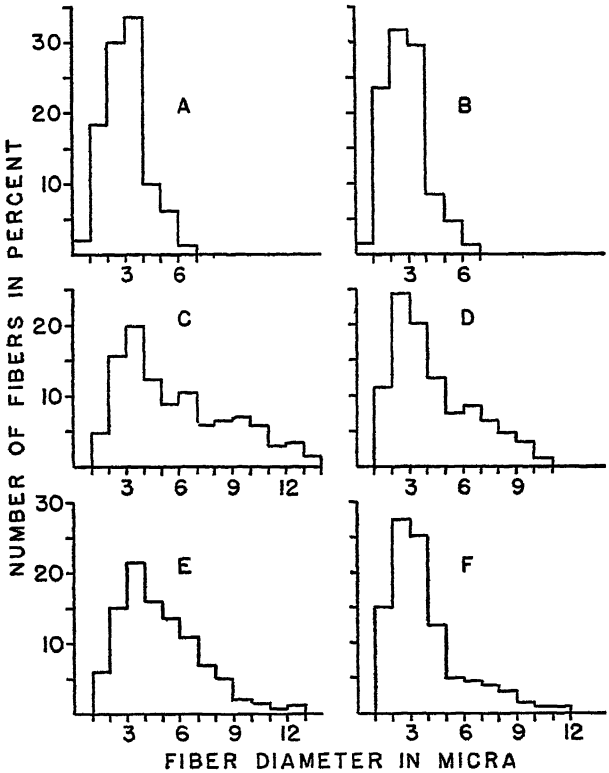


FIGURE 6 The effects of delayed suture on fiber diameter growth in the distal stumps of regenerating nerves Control sutures with no delay on the left (A, C E) · delayed sutures, on the right (B, D, F).

A is from the left tibial nerve 105 days after immediate suture; B is from the right nerve of the same cat 105 days after a suture which was delayed 84 days; C and D, 440 days after suture with 253 days delay in D, E and F 337 days after suture with 476 days delay in F.

FIGURE 6 C was from the same level of the opposite tibial, with no delay. Both nerves were allowed to regenerate for 440 days after suture, and it must be pointed out that the longer regeneration period may be an important factor. The fibers in FIGURE 6 D were smaller throughout, with a maximum diameter under 11 micra. On the control side, FIGURE 6 C, the fibers were generally larger, with a few reaching 14 micra.

The effects of delay were even more marked after 476 days, as shown in FIGURE 6, E and F. The histogram in F, after 476 days delay, shows most of the fibers to be less than 4 or 5 micra, while the histogram of

the control in E shows much larger fibers throughout (FIGURE 6). Both nerves were allowed to regenerate 337 days after suture.

The maximum delay which caused no change in fiber growth could not be determined accurately from these experiments, because the regeneration time was not kept constant. However, in those experiments with 253 and 440 days delay, approximately a year was allowed for regeneration. These experiments were more conclusive, and indicated that such delay periods restrict the diameter growth of the regenerating fibers. Not only were the largest caliber fibers limited in growth, but the whole fiber distribution curve was altered.

The Effects of Cross-Suturing Nerves of Different Fiber Caliber

On the basis of cross-suture experiments, Simpson and Young[4] described a restrictive influence on fiber diameter growth by very small Schwann tubes. Using a similar approach, Hammond and Hinsey[6] cross-sutured the hypoglossal nerve and the cervical sympathetic trunk. The choice of these nerves was fortunate, since they contain much different tube diameter distributions, and they are situated close together for easy manipulation for cross-suturing.

Fiber diameter distributions of the normal hypoglossal nerves, measured from osmic preparations, showed large fibers, between 2.3 and 17 micra in outside diameter, with a unimodal peak between 6.5 and 8.5 micra. Calculation from 11 hypoglossal nerves showed the median fiber diameter to be 7.7 micra, with only 3% of the fibers smaller than 4.5 micra. Similar observation on the cervical sympathetic trunk at its rostral end showed relatively small fibers, between 1.2 and 8.4 micra in diameter, with a unimodal peak at 2.5 to 3.5 micra. The median diameter calculated from 7 experiments was 3.3 micra, and only 7% of the fibers were larger than 4.5 micra.

In order to compare the effects of suture of a nerve with large fibers into a nerve with smaller fibers and the effects of a control experiment in which the nerve with large fibers was sutured into its own distal stump, it was necessary to run simultaneous experiments, with the nerves excised and measured at the same intervals after suture. Therefore, in one set of experiments, the hypoglossal nerves were sectioned and immediately joined to their own distal stumps. The fiber distribution in the distal stumps of these hypoglossal-to-hypoglossal sutures was determined as illustrated in the top row of histograms in FIGURE 7. The progress of diameter growth of the fibers is indicated by a shift of the curves from left to right, as the time of regeneration increased to 216, 250, 300, and 365 days. This increase is similar to that found

in the experiments on the tibial, peroneal, and saphenous nerves, and about 80% recovery of normal fiber diameter was found at 365 days.

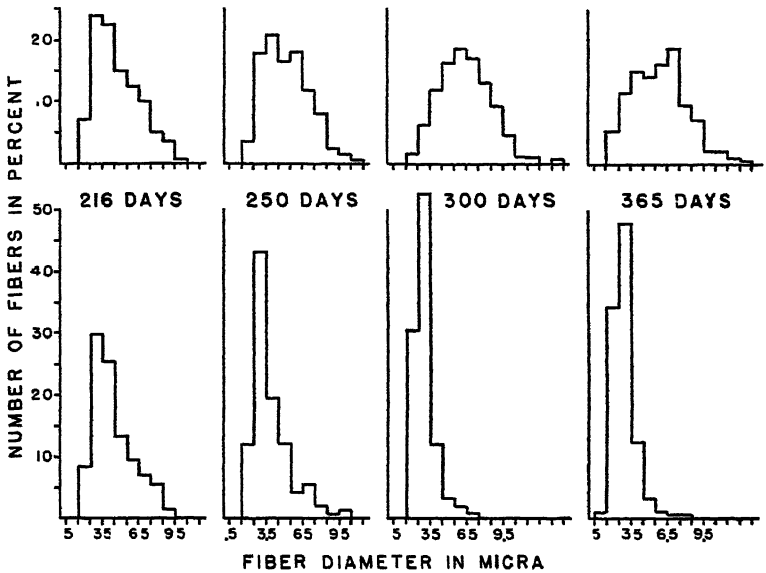


FIGURE 7 The top row of fiber distribution graphs is from distal stumps, after simple suture of the hypoglossal nerve; bottom row, from the distal stump when the central end of the hypoglossal was sutured to the cervical sympathetic trunk

The time allowed for regeneration in both the top and bottom rows is indicated in days (Modified from **Hammond & Hinsey.**[6])

With the diameter studies of the hypoglossal-to-hypoglossal series as a basis for comparison, the suture of the hypoglossal nerve to the cervical sympathetic trunk produced remarkably different results. The lower row of histograms in FIGURE 7 shows the fiber distributions in the distal stumps (cervical sympathetic), at the same intervals of regeneration as those in the upper row of hypoglossal-to-hypoglossal experiments. At 216 days, the histogram of the distal segment shown in the lower row was not much different than that of the distal segment in the upper row. However, at 250, 300, and 365 days, the fibers did not continue to grow as in the simple hypoglossal-to-hypoglossal suture, but, instead, the diameters decreased. By 365 days, the distal segment of the hypoglossal-to-cervical sympathetic cross-suture took on the diameter characteristics of the normal cervical sympathetic trunk, instead of the hypoglossal nerve.

The converse experiments, those of cross-suturing the cervical sympathetic trunk into the hypoglossal nerve, were carried out to find out if the larger Schwann tubes of the distal segment would allow the re-

generating fibers to expand beyond the diameters of their parent fibers in the cervical sympathetic trunk. The results showed no expansion, and the distal fibers tended to recover the characteristics of the cervical sympathetic trunk.

DISCUSSION

The growth of fiber diameter is part of a process of reconstitution or maturation of the fibers and can be considered separately from the longitudinal growth of the fibers toward the periphery. Although this latter process of outgrowth has been the subject of numerous studies, the diameter growth was not investigated intensively until the research of Gutmann and Sanders[1] on rabbit nerves. They described a gradual increase in fiber diameter for the first year of regeneration after suture, without recovery of either the maximum diameter or of the bimodal fiber distribution. Our experiments confirm these findings and, in addition, show that the increase in diameter continues beyond one year to at least 544 days. Also, at extremely long times, up to 1363 days after suture, complete recovery of diameter is not attained, nor is the bimodal distribution of fibers. The fact that crushing allows much more rapid reconstitution of fiber diameter confirms their findings (Gutmann and Sanders[1]).

The recovery of impulse conduction velocity and action potential characteristics is also an important part of this reconstruction or maturation process in regenerating nerve fibers. In fact, the proper coordination of nerve functions might be impossible if their conduction velocities are not regained, in spite of proper peripheral connections. These combined electrical and microscopic experiments showed that the conduction velocities and action potentials recover slowly, at the same rates, as the fiber diameters increase. The conduction velocities were actually compared to the fiber diameters of the distal stump, and the same linear relationship between these two functions was found as expected for normal nerves, as reported by Gasser and Grundfest.[5] Therefore, the electrical characteristics of the regenerating outgrowths in the distal stumps were found to be related solely to the reconstitution of fiber diameter and did not otherwise depend on the type or size of the parent fiber in the central stump. Further data on the action potentials from regenerating nerves have been reported by Berry, Grundfest, and Hinsey.[7]

A long delay between sectioning the nerve and subsequent suture was shown to impede the usual reconstitution of fiber diameter and con-

duction velocity. Holmes and Young[3] have described this phenomenon and have shown that the connective tissue tubes in the distal stump undergo shrinkage during the delay period. The experiments reported here show that delay causes even greater influence than they described. This difference in results is probably due to the fact that our experiments allowed the nerves to regenerate for much longer periods, during which the restrictive influences could be more strongly exerted. However, the results show that brief delay periods have little effect, but that delays of 253 and 476 days produced considerable interference with fiber reconstitution. Unfortunately, the exact delay times between no effect and slight effect could not be determined from these experiments. The introduction of control cross-sutures, without delay in nerves of the opposite leg of the same animals, seems to rule out the factors operating at the suture line, or differences in peripheral re-innervation (unless atrophy of the muscle is considered), which can influence the fiber reconstitution. Presumably, therefore, only differences in the condition of the connective tissue and Schwann tubes of the distal stump are responsible for the results in these experiments.

The influence of the connective tissue or Schwann tubes in the distal stump on the fiber growth has been recently emphasized by Sanders and Young,[8] who found that the motor branches of a sutured, mixed nerve contained fibers of larger caliber than the sensory branches. Also, Simpson and Young[4] cross-sutured somatic nerves into the splanchnic and anterior mesenteric nerves and suggested that the restriction in fiber diameter which resulted might be due, in part, to the small size of the peripheral tubes. The results reported here by Hammond and Hinsey[6] showed this same restriction in cross-sutures of the hypoglossal and cervical sympathetic. However, in these experiments, the nerves were allowed to regenerate for longer periods than reported by Simpson and Young,[4] and an additional phenomenon was disclosed. At 216 days after cross-suture of the hypoglossal to the cervical sympathetic, the recovery of fiber diameter was slightly less than that obtained in control, hypoglossal-to-hypoglossal, sutures. At 250, 300, and 365 days, the fibers not only showed greater restriction of growth, but actually the caliber of the fibers found distally became smaller than they were at 216 days. There is no conclusive explanation of this apparent reversal of diameter growth, but two facts might be mentioned. First, the final histogram (FIGURE 7), at 365 days, resembled that of the original distal stump before operation, and, perhaps, the small tubes compressed or killed off the larger fibers. Secondly, it must be recognized that the hypoglossal fibers could not reach proper end organs by grow-

ing down the cervical sympathetic trunk. Simpson and Young[4] showed the importance of these peripheral connections by cutting a regenerating nerve peripheral to the original suture, which prevented the re-establishment of peripheral connections. Weiss and Taylor[9] also found evidence that fibers were smaller when re-innervation of the end organs was prevented.

SUMMARY

1. Excised, distal stumps of tibial, peroneal, and saphenous nerves of cats were studied oscillographically and microscopically, at intervals, up to 1363 days after transection and suture.

2. The processes of maturation or reconstitution of fiber diameter and impulse conduction velocity continued over a long period of at least 544 days. The regenerating fibers never completely recovered.

3. Crushed nerves recovered fiber diameter and conduction velocity more rapidly than sutured nerves.

4. Delay between transection and suture of more than 6 months interfered with the reconstitution of the regenerating fibers.

5. Cross-suturing a nerve containing large fibers into a distal stump containing small connective tissue or Schwann tubes resulted in restriction of fiber diameter growth.

REFERENCES

1. **Gutmann, E., & F. K. Sanders**
1943. J. Physiol. **101**: 489.
2. **Weiss, P.**
1944. J. Neurosurg. **1**: 400.
3. **Holmes, W., & J. Z. Young**
1942. J. Anat. **77**: 63.
4. **Simpson, S. A., & J. Z. Young**
1945. J. Anat. **79**: 48.
5. **Gasser, H. S., & H. Grundfest**
1939. Am. J. Physiol. **127**: 393.
6. **Hammond, W. S., & J. C. Hinsey**
1945. J. Comp. Neurol. **83**: 79.
7. **Berry, C. M., H. Grundfest, & J. C. Hinsey**
1944. J. Neurophysiol. **7**: 103.
8. **Sanders, E. K., & J. Z. Young**
1944. J. Physiol. **103**: 119.
9. **Weiss, P., & A. C. Taylor**
1944. J. exp. Zool. **95**: 233.

PLATE 5

PLATE 5

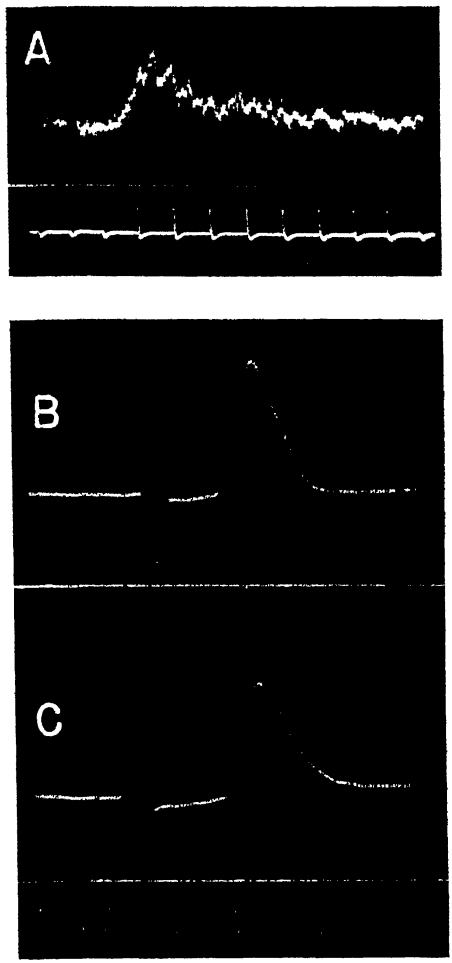
Action potentials recorded from excised distal stumps of regenerating tibial nerves.

A was recorded after 36 days of regeneration with a monopolar electrode placed 3.5 cm. distal to the suture, and the conduction distance was 3.5 cm. The spike is approximately 25 microvolts, and the time signals under A are 1.7 millisec. per cycle.

B was recorded from a normal tibial nerve with a conduction distance of 8 cm., and B is a control record for C.

C is from the opposite regenerating tibial, 1363 days after suture with the same 8 cm. conduction distance, and the same amplification as B.

The time line for B and C is 1 millisec. per cycle.



BERRY AND HINSEY: RECOVERY OF CONDUCTION IN FIBERS

NERVE METABOLISM AND FUNCTION*

A CRITIQUE OF THE ROLE OF ACETYLCHOLINE

By R. W. Gerard

Department of Physiology, The University of Chicago, Chicago, Illinois.

INTRODUCTION

Clearly, the acetylcholine system is the theme around which these papers have been arranged. The various hypotheses as to its functional significance, and especially the one regarding it as an essential component in conduction in the nerve fiber, have proven most fertile in research suggestions—witness the many studies here reported and the animated discussion of them. Yet, I must close with the judgment, on the basis of what has been said here, that this hypothesis has now exhausted its usefulness

May I first offer, as evidence of my own long sympathy to the view I shall shortly be dissecting, a quotation or two from my early writings?

"It remains to correlate this material [on heat and metabolism] with some actual mechanism of conduction. The current view that activity of one portion of a nerve fiber is the stimulus to the adjacent portion and so along the entire fiber has much to support it, especially in the form developed by Lillie. Recent evidence indicates that conduction itself may be analyzed into two phases occurring repeatedly in succession. The first is an explosive type of chemical change in a portion of the membrane surrounding the nerve fiber, and it leads, probably by local potentials, to ion movements within the fiber, which constitute the second phase. Local concentration of ions against an adjacent portion of membrane initiates here the explosive change, and so on. Probably the ion movements are associated with only a small fraction of the energy changes, and with the behavior of the membrane during and after conduction" (p. 499[1]).

" . . . In this way, it is obvious, a wave of electric and chemical change must spread along the nerve fiber in both directions from the point first stimulated. This is the nerve impulse, a propagated excitation. . . . Certain steps in this development are hypothetical, and it must be recognized that the picture has been simplified to a merest skeleton.

* This paper is essentially as presented on February 9, 1946. Later developments of any kind have not been introduced into the discussion.

The action potential, for example, may not represent a passive depolarization but a potential actively produced by the chemical reactions. But whatever the details, it is highly probable that the nerve impulse consists basically of a local membrane change of a chemical and physical nature, which leads to a flow of ions, or current, which in turn starts the local membrane change at adjacent points" (p. 64[2]).

" . . . Either the same kind of ion migration and chemical response which represents successive activation of one region of the nerve fibre by another must also take place at the synapse, or it is conceivable that the end of the axone acts as a miniature gland and, when stimulated, produces some chemical which is able to excite an adjacent or neighboring dendrite" (p. 74[2]).

"In the nervous system itself, a similar mechanism has been considered by several workers. The end of an axone is at least an unspecialized end organ, often a complicated one (as, for example, in the olfactory glomeruli) and might activate the dendrite or cell body on which it impinges via chemical as well as electrical changes. The transmission from cell to cell by means of action potentials has long been the orthodox view, and emphasis on the chemical possibilities has had a novel flavor. In fact, however, the conduction along a nerve fibre involves excitation of a resting region by an active one, and both electrical and chemical components are present in the mechanism of propagation. At the ending, which is specialized, at least anatomically, either or both components might well be exaggerated to facilitate transmission over a critical region. Long-enduring action or depolarization potentials or special chemical accumulation might equally well be utilized in various situations and (except for familiarity with the one idea) one seems as likely as the other" (p. 546[3]).

GENERAL BACKGROUND

The Role of Metabolism

Nerve fibers, like whole neurones or any other cells, depend on a maintained metabolism to survive and to function. This was strongly indicated when it was found,[4] near the start of this century, that nerve conduction failed in the absence of oxygen; and was proved when nerve respiration and heat production, at rest and on activity, were successfully measured by several workers in the mid-twenties.[3, 5] The next question is: For what result is metabolism essential? Or, What agencies link the chemical reactions with the physiological consequences? In general, the answer is clear enough: Metabolism liberates

energy to do necessary work, such as to maintain polarization across a leaky membrane, or it removes unwanted substances, or produces required ones. Further, the change in concentration of a substance may or may not be a needed step in the event of functioning, an indispensable gear in the cell machine.

In a particular case, it is often a teasing problem to determine just what role a metabolic event plays in a tissue's function. The formation of lactate in muscle contraction is a perfect illustration. When this relation was first established, early this century, lactic acid was at once assigned the key role of initiating shortening. It was an essential gear and, perhaps by changing surface tension due to acidity, engaged the shortening mechanism. Its removal or neutralization permitted relaxation. Later, attention to energy balance emphasized that glycolysis could supply the energy required in anaerobic contraction and that this reaction was largely rewound with oxidative metabolism in oxygen. It was an easy assumption, then, even under aerobic conditions when no lactate change was found, that there was a rapid formation and destruction of this substance. Indeed, lactic acid was considered the essential link between metabolism, of which it was a necessary intermediate, and contraction, of which it was a necessary cause, and it was supposedly involved in both energetics and mechanics.

Iodoacetic acid, alactic contractions, phosphocreatin and adenosine triphosphate changes, and, finally, the use of lipid fuels (not to mention myosin), changed all that.[6] Muscle did not require lactic or any other acid to shorten it; lactate is not part of the machinery. The CrP and ATP breakdown supplied the early energy needed for contraction, heat, and work; lactate formation is not an immediate energy source. Moderate exercise with good oxygenation involved no lactate change and little carbohydrate loss; lactate is in no way necessary to contraction. It is just one of the many initial or intermediate fuels available to the engine under normal working conditions, and its accumulation anaerobically is, in a sense, a sign of failure to complete the initiated oxidations.

In a particular case, further, it is well to note that historical accidents greatly influence the trend of our scientific thought and research. Acetylcholine first came to attention as a pharmacologic agent; ATP, as an intracellular substance involved in important metabolic sequences. The great experimental sweeps were, accordingly, oriented differently in the two cases. Yet ATP also has profound pharmacological actions,[7, 8, 9] and ACh may well prove to be an important component in cell metabolic systems in general. This point will require attention later.

Neural Metabolism

Finally, I shall recall the gross metabolic picture of neural tissue.[10-14] For nerve, in contrast to muscle, the heat and respiration balance of rest and action were first established. Lactic acid, next studied, seemed to be excluded from any role except as an anaerobic end-product, but it was later shown to serve as an effective substitute fuel when sugar oxidation was interfered with by iodoacetic acid; and anaerobic glycolysis was similarly established as a source of useful energy. Yet oxidations, by oxygen or an oxidizing reserve, remain of especial importance in nerve, for the long and large delayed heat production cannot be dissociated from the initial heat by anoxia or by any other maneuver tried. What fuel or fuels are oxidized, is largely unknown. At rest, nerve may destroy more carbohydrate than could be fully oxidized, while the R. Q. hovers at 0.8; and CHO utilization can taper off to zero while O_2 consumption is maintained unaltered. Even brain, with a resting R. Q. of 1.0 and a CHO fuel, can shift to another substrate which fully supports respiration. During activity, the much increased oxidative metabolism of nerve is not supported by CHO. The R. Q. of the extra respiration does rise to nearly 1.0 in tetanized nerves, but CHO loss is not increased. Some rise in acid-soluble phosphorus and in ammonia-liberation occurs, suggesting the degradation of phospholipins or phosphoproteins; but the problem is still wide open. The lipo-protein changes in rods on illumination (mentioned by Wald) and the Swedish work[15] on nucleoprotein decrease in fatigued nerve (to which Schmitt called attention), are observations challenging a resolution of this enduring uncertainty.

THE NERVE MACHINE

The resting metabolism of nerve is essential to keeping the tissue functionable. The normal resting potential, for example, falls when respiration is prevented[16] and even more rapidly when glycolysis is also blocked.[17] When an action is evoked, the cell machinery whirs, physical and chemical changes occur, an impulse is propagated, and, finally, a cycle is completed and the machine fully reset. The events associated with activity are known in moderate detail, and it will be helpful to outline this sequence. Since so much attention has been given by investigators to the early and the electrical phenomena of response, may I emphasize that all the phenomena are closely coupled together. A single impulse, gone by in a millisecond, is yet irrevocably followed by a rise in heat liberation and in oxygen consumption which

endure for minutes. Conduction fails when respiration[4, 16] or glycolysis[18] is disturbed, although, whether this is a result of interference directly with active metabolism, or is secondary to interference with resting metabolism, is not clear.[19] A mechanism, partly in terms of phosphate intermediates, for insuring the one-to-one relation between early and late events, was suggested some time ago[20] and is still useful. But, before pursuing this aspect, what of conduction itself?

Depolarization

Electric currents, applied to nerve or muscle, excite at the cathode, where ion movements are such as to depolarize the polarized membrane. The most direct evidence for the preexisting membrane potential and for its diminution by trans-membrane currents is that from impaled single nerve[21, 22] and muscle fibers.[23] Membrane potentials up to 80 mV. have been obtained from resting units; and excitation is easily achieved with a cathode outside and anode inside the fiber, but even 100-fold greater currents in the reverse direction are ineffective. Further, recalling the uniquely high sensitivity of these tissues to electric currents and the generation of electric changes when non-electric stimuli are applied, it seems probable that membrane depolarization by ion movements is the initial step in all forms of natural excitation of nerve and muscle. Whether excitation results most directly from a potential, impedance, or other, change, and to what critical level, is a separate and secondary problem.

Active Membrane Participation

There is much evidence that the nerve membrane does not passively follow the imposed depolarization, at least when applied currents are more than a few per cent of threshold, but responds with active changes. These changes are almost certainly chemical as well as physical. The decreased impedance is suggestive, but perhaps not convincing, on this point. The existence of prepotentials (with depolarizing shocks, but not with equal ones in the reverse direction) in invertebrate[24, 25] and vertebrate nerve[26] has been several times referred to in this publication. The fact that these often oscillate, and that the oscillations can increment without additional external change,[24] has been emphasized here by the report of Bronk and Brink, and by Cole's discussion. The oscillation period, 4 to 5 msec., observed in Ca-depleted nerves (Bronk), fits satisfactorily with the physical constants of the membrane, mentioned by Curtis, which should lead to resonance at about 250 cycles per sec. But such physical factors control only the period of oscilla-

tion, and cannot supply the energy to maintain, even less to increase, it. Energy, presumably liberated from metabolic events, must supply the drive, while physical conditions only modulate its flow. The analogy used in connection with the similar problem of electric oscillations in the isolated frog brain[27] may be used again. Air pressure, from the motor, drives most windshield wipers, and their beat does rise and fall with this; but the beat is much more under the control of a valve, which determines when each stroke is tripped off.

Another set of facts bespeaks, even more strongly, the intervention of a chemical step this early in the excitation process. In tortoise auricle,[28] crab nerve,[29] and even frog nerve[26] (see also Gerard,[30] and following discussion), an opposed electric shock, delivered between a supra-threshold shock and the start of the resulting propagated response, can nullify the response. While it may be possible for the purely physical changes, produced by a pulse in an appropriate network, to surge on to a peak after the pulse has passed (as Curtis suggested in connection with the reversed action potential), such a physical interpretation is under the burden of offering positive evidence in the case of the cooled auricle, where a reverse shock given 20 msec. after an effective one is still able to abort the response.

The Discontinuous Response

When the local membrane changes have progressed sufficiently, a full-fledged action appears and propagates. This response, as several speakers have emphasized, is not a continuation of the earlier processes, but a new and explosive group of events. Here, even more surely than in the preceding phase, chemical as well as physical changes are involved. The resting membrane potential shifts abruptly, not merely toward or to neutrality, but to an inverted magnitude which can much exceed the original level.[21, 22, 23] Perhaps, as Curtis suggests, this is only a physical overshoot, rather than a newly-developed, oppositely-oriented, and chemically-active membrane battery; but the burden of proof seems to be clearly on the adherents to such a physical view. Cole's comment, that the reversed action potential can vary in magnitude independently of the resting potential, certainly favors more the positive conclusion. Höber's suggestion, that a fatty acid is released by activated lecithinase and, reaching the inside of the membrane, reverses its potential, just as caproic acid does when placed on the outside, is an example of the chemical, active-membrane-change viewpoint. (This particular example is not fully satisfying, however; for, if the non-polar chains enter the membrane lipids and the polar

carboxyl groups form a negatively charged layer in the aqueous phase, this could shift the outer membrane surface charge from positive to negative, but could hardly shift the inner surface charge from negative to positive.)

The well-known high temperature coefficients of excitation also speak, though admittedly in an uncertain voice, for chemical components in the process. If elongated molecules in a loose palisade in the membrane are merely bent about, during stimulation, then they must make quite a sudden fall when a threshold is reached, and must also start a vigorous series of changes. For, whether the main chemical reactions of metabolism accelerate during, or only after, the explosive membrane response, they are locked to it in an essentially invariable sequence. And, finally, the important and complex impedance, and potential, variations which accompany or follow the spike surely indicate processes beyond simple ion movements or dielectric strains. The action potential spike represents more than a passive depolarization of a previously charged membrane. It is an active physico-chemical process, still unknown in its details.

Local Currents

Whatever the events in an activated membrane region, there remains no doubt as to the mechanism of propagation along a nerve or muscle fiber. Voltage differences between active and not-yet-active areas must lead to current flow between them and to catelectronic depolarization of the latter. That such currents are a sufficient mechanism for propagation is certain from the experiments in which the nerve impulse is made to jump a block one or two millimeters long.[31, 32, 33] Even normally, propagation is probably by similar saltations from node to node, in medullated fibers.[34, 35, 36]

Immediate Recovery

During the absolute refractory period, often under a millisecond, the membrane must at least recover toward its normal potential, impedance, and other properties, so that it is again activable. While the anodal action of the eddy currents sweeping on ahead may contribute to this restoration, this is obviously insufficient. Energy has been dissipated and must be made good from sources beyond the currents which help dissipate it. There can be no reasonable doubt that the complex of initial and immediately-subsequent recovery, with the reversing thresholds and potentials already well known,[17] is dependent on one or more of the energy-yielding metabolic reactions; perhaps on ATP breakdown.

Full Restoration

For completeness, although it is far removed from the direct problem of propagation, I mention, finally, the delayed recovery processes. We often forget that the increased respiration of activity persists a half-hour or more after a brief tetanus of nerve,[38] that the delayed heat production is similarly prolonged,[39] and that considerable after-potentials may endure for comparable periods.[40] And, as late recovery lags progressively further behind in a continuously-driven nerve, its response capacity falls to a lower equilibrium level. Irritability, velocity, chemical, thermal, and electrical response per impulse, etc., fall during a maintained tetanization.[41]

Now, with this outline of nerve action before us, I should like to consider the questions around which so much of this symposium has revolved: (1) How does excitation engage metabolism; (2) what is the role of the acetylcholine system; and (3) in what respects does junctional transmission differ from that in a fiber?

THE LINKAGE OF ACTION TO METABOLISM

Given the externally-applied stimulus energy, given even the propagated membrane response, the insistent question remains as to how one event induces the next and, especially, how chemical changes are made to follow the physical ones. This was asked by Grundfest, discussed by Green, and exemplified by Ochoa's contribution. It was considered for nerve, in some detail, a decade ago,[30] and is today being clarified in the case of muscle.[42] In muscle, the extra step of mechanical response offers both an additional problem, of how the membrane response leads to the myosin response, and an additional line of attack on the general case. Since metabolic details are far more numerous for muscle than for nerve, I shall choose illustrations freely also from the former material, in confidence that the principles they illustrate are equally valid for both tissues.

Ion Action

In the homogeneous liquid phase of a heterogeneous system like tissue, electric currents, applied as external stimuli or generated in the course of the active response, mean ion movements and only ion movements. Where these ion streams encounter interfaces—membranes, micelles, molecular palisades, etc.—ions can accumulate or decrease. A local change in ion concentration at molecular or structural surfaces of a cell must be the initial consequence of an electric stimulus and ap-

pears to be the only possible first link to a metabolic chain of events. (Electron shifts within single molecules or lattices, if such occur, would themselves follow the shift of charged ions, unless large electromagnetic fields were applied—fields that seem beyond a range of possible biological significance.) However, changed ion concentration is easily sufficient to initiate other local chemical changes.

Altered metabolism means changes in the rates of chemical reactions. Not only are quantitative increases or decreases of total metabolism the sum of similar variations in the rates of the component reactions; but also qualitative changes are the resultant of increase in rates of certain reactions and decreases in others. The rate of a given reaction is determined by the concentration of active reactants and products and by the catalytic conditions (temperature, water, ions, etc.), especially by the enzyme activity. Reactant concentrations can change only as a result of an antecedent change in another chemical system which produces them—which gets us no further in our problem—or of a spatial redistribution. If reactants are themselves ions, and so moved by the stimulating currents, this could be a direct result of stimulation. In most cases, however, as emphasized by Höber, such a redistribution would also demand prior changes in the system to increase the physical availability, changes in membrane barriers, surface adsorption, and the like. These might also be a direct effect of the stimulus, as the rotation of a polar molecule, but are more likely to be secondary to more extensive chemical changes.

In contrast to the relatively unpromising situation for substrate alteration, a modification of enzyme activity, and so of metabolism, by ion changes is both theoretically probable and experimentally established. Besides a direct ion effect on the activity of given enzyme molecules, there exist the other possibilities of activating pro-enzymes (Ca on prothrombin), removing inhibitors (phosphate or citrate binding calcium), and adding accelerators (Cu on thiol oxidation[43]). Such ion effects are richly present in biological systems, as well as in the non-living systems mentioned by Dr. Alexander, and it may be useful to itemize some that are important in muscle and nerve tissues.[12, 27]

Magnesium ion either is essential to, or materially hastens. a number of key reactions in carbohydrate degradation, while local increase in its concentration would suffice to initiate or accelerate them. The phosphorylation of glucose to hexose-6-phosphate by hexokinase, a reaction of especial importance in neural tissue which "prefers" glucose to glycogen as a fuel, requires Mg^{++}; as does, also, the shift of phosphate from the 1 to 6 position by glucophosphomutase.[44] The further phosphory-

lation of this substance, an intermediate in the glycogen as well as glucose reaction chain, to fructose 1,6-diphosphate, also requires Mg^{++}. The later change in phosphoglyceric acid to the energy-rich phosphoenol-pyruvic acid involves a magnesium combination with enolase, and fluoride inhibition of this reaction depends on displacement of Mg^{++} by a fluoride complex.[45] Magnesium or manganese, as well as calcium (K inhibits), is an essential component of the system which forms ATP and pyruvic acid from the phosphopyruvic acid and a lower adenosin phosphate.[46] This ion is again reported necessary for the splitting of ATP to ADP by myosin,[47] although most workers[48, 49, 50] find Mg^{++} inhibitory here. Mg^{++} also inhibits the shift from 3-phosphoglyceric acid to 2-phosphoglyceric acid by phosphotriose mutase.[51]

Calcium ion, as mentioned, is required for the formation of ATP from phosphopyruvic acid. It is also involved in splitting a phosphate from ATP by myosin or other ATP-ases[47, 52, 53] and, perhaps, in the accompanying shortening of the myosin fibers. The splitting of acetylphosphate to acetate and phosphate is accelerated by Ca^{++}.[54] And, again, to mention an inhibitory action, Ca^{++} interferes with the formation of acetylcholine from choline.[55]

The formation of ATP is thus influenced by all three of the major cellular cations (Mg, Ca, K); its destruction, by at least two (Mg, Ca). ATP, in turn, is critical in both fat and carbohydrate oxidation and may be one of the regulators of metabolism. Thus, a lowered ATP concentration might favor utilization of carbohydrate over fat[56] and glycolysis over respiration.[57] The abrupt shift of muscle metabolism, on vigorous contraction, in just these directions, may, then, be due to the fall in ATP concentration and this, in turn, to the movement of ions to or from the critical enzyme surfaces.

Another example of ion importance, especially of K^+, is offered by recent myosin studies. In an appropriate system, myosin B contracts in 0.1 M KCl and relaxes when the K^+ is doubled in concentration[58]; and an antagonism between K^+ and Ca^{++}, or Mg^{++} and Ca^{++}, on myosin action as ATP-ase and on myosin extension, has been repeatedly noted. Potassium also increases the content of creatin phosphate in muscle (while Ca^{++} decreases it),[59, 60, 10] as well as of stable phosphate esters in nerve,[61] perhaps by its ability to enhance CrP formation when pyruvic acid is oxidized.[62] This ion also aids both the synthesis and the liberation of ACh and, conversely, ACh (or ATP) can release K^+.[63, 55, 64]

It is not difficult to trace connections between these catalytic actions of the tissue cations and the physiologic effects which ion changes pro-

duce; but, at present, relating specific actions to specific effects would be mainly guesswork. It will suffice to recall that K^+ increase leads to such effects as: a rise, passing into a severe fall, for the irritability, membrane potential, and electrotonic spread in nerve fibers, and, perhaps, for their oxygen consumption, as Brink mentioned; and a fall, from the start, for the spike (only slight), after-potential, conduction velocity, and recovery rate. Similar changes have been observed less fully in muscle and in the central nervous system: moderately increased K^+, for example,[12, 27, 65] increases the fast electrical activity of cat or frog brain and prolongs the after-discharge on stimulation of deep cerebellar nuclei (see also [66]). There are, thus, ample roads from current flow, through altered ion concentration and chemical reaction rates, to physiological responses. The problem is not to find connections, but rather to identify the few important actualities among the many conceivable possibilities This brings us back to acetylcholine.

THE ROLE OF THE ACETYLCHOLINE SYSTEM

That ACh is formed and destroyed as an integral part of impulse propagation in nerve fibers, has been suggested by several workers[67, 68] and strongly supported by Nachmansohn.[69] He has summarized his arguments here: ACh is present in nerve and is released on stimulation (though its leaving the cell is accidental); cholinesterase (ChE) in nerve is highly active and specific, and choline acetylase (ChA) is also rich in nerve; a close parallelism exists, in the electric organ, between potential and ChE activity; energy relations, in electric organ and nerve, are satisfactory for ACh synthesis via CrP, etc.; various drug actions, though always in danger of misinterpretation, especially where penetration through a membrane is involved, do support the importance of ACh in conduction. Just what the role of ACh is, seems less defined. Earlier, Nachmansohn supposed that the stimulus liberated ACh directly and that this caused the membrane depolarization. Now, recognizing that the stimulus itself must lead to depolarization, he suggests that ACh is responsible for the loss of resistance in the membrane—certainly, a step for which a chemical mechanism would be welcome. Beutner and Barnes have also emphasized a function for ACh, both in producing the action potential and in lowering membrane resistance.

Quantitative Relations

The calculations (Nachmansohn), that the ChE at a motor endplate is powerful enough to split a complete layer of ACh in a milli-

second (perhaps even in a few microseconds, which Cole has mentioned as a more appropriate time for the impedance changes in nerve), and that the enzyme activity in the C.N.S. would split a layer covering 10^8 sq. mu. per gm, are impressive. There is no question of strong esterase activity. However, it is worth noting, for the C.N.S., that even the surface of the nuclei in a gram of brain, let alone the whole neuronal surface, is some 4×10^{10} sq. mu.[70] The difficulties appear when one examines the rest of the system, to see how well it can keep up with the esterase. Let us assume, with Nachmansohn, that some 2000 cal/M are required to esterify choline to ACh, and calculate, from his data on enzyme content, Feldberg's summary[71] of ACh content and liberation, and the figures of myself and others on heat and metabolism, the over-all balance for nerve and brain.

Nachmansohn estimates that, in mammalian brain, ChE can split up to 10^{15} molecules of ACh per millisecond per gram fresh tissue. This amounts to 6 millimoles per hour per gram. In terms of Q_{CIC} values, reduced to these same units (mM, hr, gm.), less ACh could be split: between 0 3 mM for cortex, and 3.0 for caudate nucleus or sympathetic ganglia. For mammalian nerve, the rate calculates to 0 06; for frog nerve (20°), to 0.05; and for white matter, to 0.02. In contrast, the maximum rate of ACh synthesis (in tissue brei in N_2, with ATP and all necessary accessories) is 0.001 mM for mammalian brain and 0.0005 for nerve, still in these same units. In both mammalian brain and nerve, therefore, ChE activity is over 1000 times, perhaps over 5000 times, as great as ChA activity, and similar relations will probably be found for the frog. Neural enzymes can split ACh by three or four magnitudes faster than they can build it.

This calculation is made, of course, for maximum rates and over long time intervals, and requires further consideration. If, for example, the synthesis normally continues evenly in time, but the hydrolysis occurs only in brief bursts associated with activity, the discrepancy in rates might be unimportant. But this will not hold. First, whether ACh be associated with the potential or impedance changes of a nerve action, the rise is far more rapid than the fall, and the need for an explosive release of the agent is even more imperative than for an explosive destruction. If, therefore, ACh is synthesized and destroyed in the course of each nerve action, ChA should actually be several-fold more active than ChE, instead of a thousand-fold less active. Let us make the more favorable assumption, however, that ACh need not actually be synthesized for each impulse, but only be released from a store. Then, though used in bursts, its formation could be continuous. Even so, there remain fatal discrepancies.

The total ACh present in whole brain is, keeping to millimoles per gram fresh tissue, about 2×10^{-5} for mammals and twice as much for the frog. White matter contains ten-fold less, but the value for mammalian mixed nerve is close to 3×10^{-5}. Dorsal roots contain, at most, one-twentieth of this amount; frog nerve,[67] even less, 10^{-6}. All the ACh in mammalian brain could, therefore, be destroyed by ChE in about 50 milliseconds, and would require a minute to be synthesized by ChA. For mammalian nerve, the stored ACh could last less than two seconds and would require over three minutes to replace; for frog nerve, the ACh could last about 65 milliseconds, and for mammalian white matter (not to mention dorsal root), a third of a second. Yet nerve, including roots and central tracts, can maintain activity for hours, conducting hundreds of impulses per second; and activity of the central grey can also long outlast the possible time limits. Clearly, then, neither ACh storage nor synthesis, nor both combined, could possibly (unless an entirely different order of ChA exists *in vivo* than has been found in extracts) supply this substrate as fast as ChE can split it.

Of course, an enzyme is not always kept saturated with substrate. However, this at once undermines the many arguments that have been made, from high ChE concentration, for the possibility of rapid rise and fall of ACh concentration; it also throws into question the significance of high local ChE concentrations. The striking finding, for example, that ChE is 15,000 times or more as concentrated in the end-plate region as in the adjoining nerve or muscle, adds confusion rather than insight. The end-plate potential falls much less rapidly than that of nerve or muscle, and there is no evidence of a great store or synthesis of ACh there. How, then, can the tremendous ChE activity be reconciled with any current theories relating ACh to neural functioning? When a 2000 horse-power engine is found in a half-ton truck, one must suspect it is there for some other reason than to supply ordinary motive power.

Let us agree, however, that ChE is not kept fully saturated, and continue with these calculations. The cat cervical sympathetic ganglion releases ACh to perfusing fluid, on preganglionic stimulation. Again in millimoles per gram, the ACh content of the ganglion, before or after several hours' tetanus, is about 10^{-5}, although five times this amount has been released during the activity period. The rate of release falls with continued activity, but holds up better when some blood is present. A maximum of 10^{-4} is liberated in five minutes' tetanus at 17 per second or, per impulse, about 2×10^{-8} (cf. [72, 73]). The ACh es-

caping from the ganglion, per impulse, is thus only about one-fiftieth of the amount that the ChE present could split in one millisecond. If ChA activity is taken to be one-thousandth that of ChE, the ACh synthesized in 60 ms. (the interval between impulses) could equal that released or exceed it two- or three-fold. This, incidentally, leaves no place for the often-assumed existence of a much greater ACh turnover within active units than is reflected in the amount escaping from them.

Isolated frog nerve, according to von Muralt,[67] actually increases its ACh content on tetanization, from 10^{-6} at rest to 1.5×10^{-6} while active, and the ACh increase per impulse calculates to 6×10^{-8} from von Muralt's figures, to 10^{-11} from Lissak's.[68] (In the latter experiments, only the ACh diffusing from the cut ends of a stimulated nerve was measured.) For cat gastrocnemius, assuming a weight of 20 grams, the ACh released by a single maximal twitch evoked by the nerve is 6×10^{-12} mM/gm.[74] Nerve can, of course, conduct several hundred impulses per second for long periods, but we might conservatively calculate with 50 per second, or 20 milliseconds total time available per impulse. Frog nerve ChE could split in this period 3×10^{-7} mM/gm. of ACh: five times the amount von Muralt finds liberated and 30,000 times Lissak's figure. If, again, ChA is only one-thousandth as active, it could easily supply ACh at the rate demanded by Lissak but would fall short of von Muralt's figure by 100-fold. On the basis of such an analysis, a nerve should be able to conduct an impulse only once in two seconds. Von Muralt's value, incidentally, is far more in accord with that for the ganglion, both in absolute amount and in relation to ChE activity, and it is also more probably correct on methodological grounds. But it cannot be right if the assumed ChA activity is remotely correct.

Perhaps, then, all these discrepancies result from falsely low ChA values. This enzyme system might easily have been seriously injured during tissue extraction and, thus, be far more active *in vivo*. Let us make this assumption, and allow a ChA activity sufficient to equal ChE activity or, giving ACh the most favorable conditions, an activity sufficient only to cover the ACh actually released on stimulation. Note, however, that even this excludes any greater ACh formation and subsequent destruction, within the cell or outside it, beyond the measured formation. If this greater turnover is allowed, by assuming ChA activity to equal ChE, the following relationships reveal still more intolerable discrepancies.

The formation of one millimole of ACh requires, we have agreed, some 2 calories. The sympathetic ganglion, releasing ACh on stimula-

tion at the rate of 10^{-3} mM/gm./hr., would liberate 2×10^{-5} cal./gm./hr. For frog nerve, at 50 impulses per second (and at this frequency the energy per impulse is fully 80% of that at zero frequency), von Muralt's value gives 10^{-2} mM ACh or 0.02 cal. (If ChE were working at full capacity, the heat liberated just from ACh splitting would be 6 cal. for the ganglion, 0.1 cal. for frog nerve!) But the measured total heat production of frog nerve is, in these units, 0.1 cal. at rest and 0.18 at maximal activity; for mammalian cortex (using the highest values of Q_{O_2} reported[75]), the resting energy release is 25 cal. and that of maximum activity perhaps 50 cal. These brain values are probably much too high for the ganglion (probably three-fold[76a]), but this gives every advantage to ACh. The actual ACh released in nerve during activity would thus, during its normal hydrolysis by ChE, account for over 10% of the total heat of nerve activity. Yet, only 3% of this heat is initial heat, immediately related to the events of conduction. Moreover, other exothermic reactions are surely involved, even with ACh itself—in its formation, liberation, neutralization, etc.—before that of its destruction. (And again, if ChE were fully active, the ACh hydrolysis heat alone would account for more than the full extra heat production of active nerve!)

An examination, further, of actual chemical reactions involved in the synthesis of ACh raises added difficulties. The initial energy source for ACh synthesis is considered to be CrP. During maximal frog nerve activity, less than 13 mgm. % of CrP is split in an hour; enough to account, at best, for 0.007 cal.[10], far below the needs for ACh. But, of course, CrP is resynthesized by energy from other metabolic reactions, so this does not mean too much. The total fuel turnover, however, does set an inescapable limit. For bullfrog nerve, 6 mgm. % of carbohydrate disappears per gram per hour at rest or activity;[76] for the small frog nerve, this might be 10 mgm. %, or 6×10^{-4} mM/gm./hr. On complete oxidation, this could yield a maximum of 0.02 mM of CrP, if all the energy available to form high-energy phosphate bands (3 per atom of oxygen) were so directed. Thus, the total nerve metabolism could just comfortably synthesize ACh at the rate it is reported actually to form during activity (.01 mM ACh from .02 CrP), and could not begin to supply energy to synthesize it at the rate ChE can destroy it. (Actually, the picture is worse than here presented, because the maximum heat of activity is 0.18 cal./gm./hr. for frog nerve, whereas the assumed carbohydrate oxidation would yield 0.4.)

It may also deserve thought that, while the esterase is located in the membrane of the giant nerve fiber, the oxidizing enzyme systems are

distributed through its volume. It seems impossible that any considerable fraction of the oxidative energy released in the core of such a fiber could be utilized to drive reactions in its surface, up to 400 micra distant; and it seems unlikely even in the usual medullated fibers, up to 20 micra in diameter. Along the same lines, I know of no evidence for the intensive respiration at neuro-myal junctions which would be demanded to keep up with the terrific ChE activity. There is definite evidence against an intense respiration at synaptic regions in the central nervous system, despite their greater relative surfaces for fiber propagation and junctional transmission.[70] (A high ChE and DPN concentration in the synaptic layers of the retina, however, has recently been reported.[77])

A final calculation, dealing with materials rather than energetics, is only suggestive. Most workers have tacitly or explicitly assumed that the acetate formed by hydrolysis of ACh was not re-utilized for ACh synthesis. Indeed, ACh has been found by Lipton[78] to form only from pyruvate in oxygen or an acetate source (acetoacetic or citric acids) in nitrogen, under present *in vitro* conditions. This would exclude full re-utilization of acetate, even in nitrogen, unless the reaction, 2 acetate → acetoacetate, is fully reversible. Lipmann has just indicated that synthesis from acetate may be possible when additional components, of coenzyme character, are added, and this would make easier a cyclic use of acetate. Without cyclic use, a molecule of glucose would have to be lost for every two of ACh formed and hydrolyzed—or twenty-fold the actual rate in nerve, 0.01 mM ACh; 0.0006 glucose. The accompanying heat production would have to be similarly outrageous, in comparison with the factual rate.

Such quantitative considerations are admittedly rough, with little attention to detailed conditions (temperature, species, rate of stimulation, etc.), but the order of magnitude cannot be far off. They demonstrate conclusively, I believe, that ChE cannot possibly exert its full activity on ACh in neural tissues and suggest that other meanings for its presence and action be sought. Further, even the less drastic rates and amounts reported for other phases of an ACh system lead to severe quantitative strains on the total metabolism of nerve or brain. But still other difficulties have been brought out in this publication.

Drug Action

Drugs, especially esterase inhibitors, have been widely used in studying the ACh system, and their actions have been much discussed during this symposium. The point of greatest debate has been the ques-

tion of permeability; for, of course, the absence of an expected effect *in vivo* could easily be due to a failure of the added substance to penetrate to the vulnerable region. However, the evidence marshalled seems to be conclusive that esterase can be inactivated, or ACh content increased, without serious disturbance of function of nerve or muscle.

All agree that eserine, a tertiary amine, can enter nerve and muscle, and Nachmansohn makes the point that the action potential of squid nerve can be abolished by soaking in this drug. (The fall of the action potential does not show the great prolongation one might expect if ACh removal were interfered with.) Yet Cantoni and Loewi have reported[79] that a frog can be eserinized *in vivo* so that nerve ChE activity is abolished, while nerve conduction remains undisturbed. (Conversely, intravenous ChE blocks the pupillary reflex in rats.[80]) A comparable result with the even more powerful, and irreversible, inhibitor, diisopropyl-fluorophosphate, has just been presented by Gilman and by Bodansky. Both *in vivo* and *in vitro*, though with some anomalies in behavior, this agent has been shown to inactivate entirely ChE while leaving nerve conduction and action potentials intact. Although detailed criticisms have been made, especially by Talbot, the major fact remains, as in the eserine experiments, that conduction without esterase is possible.* Again, veratrin can inhibit ChE,[81] yet it does not[82] influence muscle or the neuromyal junction, including its sensitivity to added ACh, except for a late and independent negativity.

The inability of ACh, added in large concentration to the surrounding medium, to depolarize nerve or otherwise to disturb conduction, has been reemphasized by the new experiments of Lorente de Nó and of Bronk. Nachmansohn has urged that ACh, a quarternary ion, cannot penetrate the lipoid membrane of nerve fibers, except at their naked terminals, thus accounting for these negative results. Yet ACh does leave nerve trunks on stimulation and should, similarly, be able to enter under combined anoxia and stimulation. Further, both Bronk and Atcheson have presented clear evidence that tetraethylammonium

* The results of Gilman and the Edgewood workers have since been challenged by Nachmansohn and his colleagues. Both groups reported work at the April meeting of the Federation, and their full papers have since appeared (J. Neurophysiol. June). Work done in the interval in my laboratory fully supports the conclusions of the Edgewood group.

Frog sciatics were immersed in peanut oil, with or without DFP, resting on stimulating and lead-off electrodes. Action potentials fail in a few minutes or remain normal for hours, depending on the drug concentration. Conduction, when lost, is not restored in fresh oil. A nerve exposed for an hour to a non-depressing concentration of DFP, washed, ground, and assayed for cholinesterase by its rate of destruction of added acetylcholine (tested on the frog's rectus), shows no cholinesterase activity. A companion nerve continues to conduct well, while remaining in the same DFP solution. Further, when a washed, poisoned nerve is ground together with an untreated one, the homogenate assays at the average cholinesterase activity of the two nerves taken separately. The DFP inactivation of cholinesterase occurs, therefore, prior to the grinding. Clearly, conduction is possible in nerve lacking cholinesterase.

chloride, another quarternary, acts powerfully on medullated nerve and so, presumably, penetrates easily.

A seemingly crucial test has been carried out in the last few weeks in our laboratory by Miss Graham. ACh Br (1:1000 in isotonic KCl with vital red) was injected into single muscle fibers of the eserinized frog sartorius with a micropipette, and the membrane potential measured. With one electrode inside the fiber and another outside, membrane potentials of 40 to 80 mV are regularly obtained. Injection of a large drop of isotonic KCl with or without ACh, large enough to fill the fiber cross-section and spread one to three diameters along its length, will immediately lower the potential by one- to two-thirds; but a smaller drop, not filling the entire cross-section, has little effect; only 7% fall, in one fully satisfactory experiment with ACh. It seems, then, that ACh inside the membrane does not depolarize it, as postulated by Nachmansohn and by Beutner and Barnes.

A number of other points have been raised here, which must also be kept in mind. Bodansky has emphasized the existence of a family of esterases, even in different neural structures of the same species: *e.g.*, the enzymes in cervical sympathetic ganglion and in brain show different substrate and ion concentration optima, different equations relating concentration to activity, different substrate selectivities, etc. The last point is especially important, since a criterion urged for discriminating between "true" and "pseudo" esterase is the relative inactivity of the true enzyme with tributyrin. Yet the "true" esterase of brain splits triacetin up to six times faster than it does ACh.

The distribution of ChE or ACh, or both, in various organs and tissues has also been mentioned by several discussants. Rosenblueth asked about conduction in adrenergic nerve fibers, which lack the ACh system; Hoagland made a similar point about Nitella, which conducts independently of ACh; another discussant mentioned a recent report that ChE is absent in the electric organ of Malaptorurus; and the carotid body, although specifically sensitive to ACh, is reported to lack ChE.[83] ChE is also absent from the iris sphincter of the amphibian eye, while present in its cornea and in the turtle's sphincter.[84] Conversely, parts of the ACh system are richly present in spleen, placenta, cornea, potatoes, and some bacteria, where any relation to neural function is nearly, or quite, impossible. And finally, in this connection, many other agents act on, and other enzymes are present in, neurones. Adrenalin keeps up the action potential in isolated cat nerve;[85] ATP stimulates smooth, as well as striped, muscle;[86] carbonic anhydrase is interestingly distributed in the brain;[87] CO_2 has marked and differential

actions on the nervous system; and so on. Thiamin is reported[88-91] to affect ACh action and synthesis and to be liberated from precursor in relatively large amounts from stimulated frog nerve even to be *the* transmitter. Cocarboxylase, like ChE, is concentrated in the nerve membrane.[92] (For discussion of further recent evidence, see Gerard and Libet.[14]) I do not see how we can reasonably select the ACh system from all this welter and just assign to it an essential role in conduction of the nerve impulse.

THE PROBLEM OF JUNCTIONAL TRANSMISSION

In the time available, the problem of junctional transmission, presented mainly by Eccles, can only be touched upon, and even so the case of autonomic effectors, the classical one of neurohumoral action, which has not been before us, will be omitted. As for the neuro-myal junction, the unquestioned facts, that ChE is more concentrated there than elsewhere in the muscle fiber by a factor of 10^4 (Nachmansohn), and that this region is more sensitive to added ACh by a similar factor (Kuffler), are impressive; along with the potentiating and prolonging action of eserine, long known for junction as well as ganglion. I am, personally, less convinced of a transmitter role of ACh at the junction than I was a few years back, but do not consider that the evidence is crucial in either direction. (The observation[93] that the lizard muscle fiber can respond to nerve stimulation at a time when ACh applied to the end-plate is ineffective, although such ACh does cause contraction when first administered, has not been explained. Also, the end-plate potential, often supposed to be set up by ACh liberation, is present in invertebrate, as in vertebrate, muscle; but the end-plate region in the former is *not* sensitive to ACh or to curare.[94]) Discussion here has been mostly on central synapses, and the reader should consider these.

As elaborated by Eccles, the currents that flow between an active fiber region and an inactive one, whether in the same or another unit, do account for the usual activation phenomena. The results of Arvanitaki[24] and of many other recent experimenters[95, 96, 97] show that threshold changes and transmission from unit to unit in simple systems are accurately and quantitatively explicable in terms of the measured currents and the known geometry. Whether Eccles' detailed analysis of the situation at a synapse will hold up as well with time as he was able to defend it here, we do not know, but there is every reason to push such thinking further. (Some difficulties are: the very variable structures which are found in synapses, where the two units may meet as

parallel fibers, spirals of one on another, multiple contacts, etc., as well as the orthodox small end-foot stuck on a large surface, like a match stick on a cheese; the problem of multiple synapses on a cell body and the relative area of end-feet and their surround; the presence of irreciprocal conduction with protoplasmic continuity, as in an asymmetrically compressed sartorius muscle. Neither these, nor the spatial theory of inhibition, nor the need for regarding the E. E. G. as an oscillating somatic potential, problems which have received attention in this publication and elsewhere,[98] can be here expanded.)

The evidence for a transmitter role of ACh in the central nervous system, on the contrary, is inferential and conflicting. Those who have read Feldberg's recent review of this question[71] must have been impressed by the poor case that can be made. Added eserine, or ACh, or both, may increase the activity of a brain region, or depress it, or cause negligible change. The failure of ACh to alter frog cord reflexes, mentioned by Eccles, is a case in point. Or ACh may excite, while eserine depresses. Atropine, on the whole, does nothing. The two compounds mentioned by Gilman, both powerful anti-esterases and both able to produce convulsions, one of which is completely antidoted by atropine, while the other is uninfluenced by it, afford an instance of the conflicting facts in this area. Strychnine is supposed to exert its action by blocking ChE, yet Tobias[99] has found the ACh content of frog and rat brains and cords decreased, if anything, by strychnine. Nembutal, conversely, increases the ACh content, although, as Bodansky mentioned, it also lowers ChE activity. If ACh is an agent for evoking neurone activity, it should increase the oxygen consumption of brain. Lipton has recently made Qo_2 measurements on rat brain slices, at my request, and found no influence of eserine (10^{-5}) alone with eserine and ACh (10^{-5}), at most a better maintenance of the usual initial rate. Incidentally, the only other observations I have found on the influence of ACh on respiration are one[100] showing an increase in salivary gland oxygen consumption, and a forgotten one from my own laboratory[101] showing a marked decrease in the oxygen consumption of nerve (uneserinized).

The best basis for invoking chemicals in synaptic transmission is that synaptic potentials, like those of the end-plate, may last much longer than could any reasonable physical discharge period for membranes with capacitances and resistances in the range known. Then one invokes some active depolarization process, as for nerve; and then this must be explained, by a chemical reaction of some sort. To be sure, chemical activity is involved, as in nerve, and quantitatively more in-

tense. But is the chemical ACh? *Quien sabe?* It should be recalled that undrugged nerve also has an enduring after-potential, which can increase in intensity for several minutes and persist for ten or more.[40]

CONCLUSION

Dr. Nachmansohn skillfully and generously organized the extraordinarily successful conference of which this is the result, to bring forth much current evidence and a full range of judgments bearing on the significance of ACh, as well as of electrical changes, for the functioning of nerve and other tissues. With these facts and arguments before us, we must conclude that ACh is not critically involved in nerve conduction, and we must be reserved in assigning it a role in junctional transmission, particularly within the nervous system. This is progress and should lead to greater progress. Our thinking and our consequent experimentation now can be directed along new lines.

This is not to say that the hypotheses which must be relinquished have been worthless, nor that the ACh system is unimportant. Hypotheses are not true or false (who can assert absolute truth?); they are useful or useless. They do or do not suggest investigations which reveal new facts, facts which discriminate between alternate views or which fill in gaps of felt ignorance or which suggest new interpretations and experiments. By such standards, the various ACh hypotheses have been good; they have been abundantly fruitful. But this fruit is ripe, and it is time for the seed of a new idea to be germinated. Fresh fruit will then ripen with time and the present crop not be husbanded until it rots or dries up.

What a new and fertile approach may be, I do not know. ACh and the enzymes that operate in the system can hardly be present adventitiously. Nature no more evolved the ACh system to mislead biochemists than it evolved the giant nerve fiber to aid physiologists. ACh has some significance to cells. Perhaps this system is a fragment of a universally important metabolic mechanism, dealing, if one must hazard a particular guess, with the manipulation of lipid molecules. Such facts or statements as the following may serve as clues. ACh prevents the splitting of CrP by muscle juice;[102] choline lack increases the turnover of phospholipids;[103] ACh can replace Ca in enabling myosin to split ATP;[52] ACh is the only system able to capture energy via both respiratory and glycolytic reactions,[53] and so is related to both respiration and carbohydrate utilization rates.[104, 105] It would still be possible for evolution to have selected this fragment of a more general system for special emphasis and functioning in particular situations; to serve, for

example, as a transmitter at parasympathetic endings. After all, ACh *is* an ion with rather striking physico-chemical properties. In just such fashion have the ubiquitous respiratory hemins of cells been selected for the special function of transporting oxygen. The parallel evolution of hemoglobin in utterly separate phyla, as the vertebrates and annelids, was a great mystery before the discovery that such related molecules as cytochrome are almost universally present in cells.

I am suggesting, then, that ACh may extend further and have more importance in cell functioning than has yet been seriously considered and that any particular role it plays in transmission is a secondary and derivative one. In arguing, as I have, for renouncing the belief that ACh has any direct function in nerve conduction and in transmission at many junctions, I am inviting those who work with the ACh system to emerge from the chrysalis which they have outgrown and to seek fresher and greater fields of intellectual nourishment.

BIBLIOGRAPHY

1. **Gerard, R. W.**
 1927. Science. **66**: 495.
2. **Gerard, R. W.**
 1931. Quart. Rev. Biol. **6**: 59.
3. **Gerard, R. W.**
 1932. Physiol. Rev. **12**: 469.
4. **Fillie, H.**
 1908. Zeitsch. f. d. Allgem. Physiol. **8**: 492.
5. **Feng, T. P.**
 1936. Ergeb. Physiol. **38**: 73.
6. **Wilhelmi, A. E.**
 1946. In: Fulton, Howell's Textbook of Physiology. See p. 56. Saunders. Philadelphia.
7. **Drury, A. N.**
 1932. J. Physiol. **74**: 147.
8. **Green, H. N., & H. B. Storer**
 1944. Brit. J. Exp. Path. **25**: 150.
9. **Buchtal, F., A. Deutsch, & G. G. Knappers**
 1944. Nature. **153**: 774.
10. **Gerard, R. W., & N. Tupikova**
 1938. J. Cell. Comp. Physiol. **12**: 325.
11. **Quastel, J. H.**
 1939. Physiol. Rev. **19**: 135.
12. **Gerard, R. W.**
 1937. Ann. Rev. Biochem. **6**: 419.
13. **Gerard, R. W.**
 1938. Arch. Neurol. & Psych. (Chicago). **40**: 985.
14. **Gerard, R. W., & B. Libet**
 1946. Progress in Neuropsychology and Psychiatry. New York.
15. **Hyden, H.**
 1943. Acta Physiol. Scand. **6**: Suppl. 17.

16. **Gerard, R. W.**
1930. J. Physiol. **92:** 498.
17. **Shanes, A. M., & D. E. S. Brown**
1942. J. Cell. Comp. Physiol. **19:** 1.
18. **Ronzoni, E.**
1931. J. Biol. Chem. **92:** iii.
19. **Chang, T. H., & R. W. Gerard**
1933. J. Physiol. **104:** 291.
20. **Gerard, R. W.**
1934. Science. **79:** Suppl. 20.
21. **Curtis, H. J., & K. S. Cole**
1940. J. Cell. Comp. Physiol. **15:** 147.
22. **Hodgekin, A. L., & A. F. Huxley**
1945. J. Physiol. **104:** 176.
23. **Graham, J., G. R. Carlson, & R. W. Gerard**
1942. Fed. Proc. **1:** 29.
24. **Arvanitaki, A. J.**
1942. J. Neurophys. **5:** 89.
25. **Hodgekin, A. L.**
1938. Proc. Roy. Soc. London B **126:** 87.
26. **Katz, B.**
1939. Electrical Excitation of Nerve. Oxford Univ. Press. London.
27. **Libet, B., & R. W. Gerard**
1939. J. Neurophys. **2:** 153.
28. **Colle, J.**
1934. Compt. rend. Soc. biol. **117:** 99.
29. **Gerard, R. W.**
1934. J. Physiol. **83:** 24.
30. **Gerard, R. W.**
1936. Cold Spring Harbor Symposia. **8:** 194.
31. **Tasaki, I.**
1939. J. Physiol. **127:** 211.
32. **Hodgekin, A. L.**
1937. J. Physiol. **90:** 183, 211.
33. **Blair, E. A., & J. Erlanger**
1939. J. Physiol. **126:** 97.
34. **Lillie, R. S.**
1925. J. Gen. Physiol. **7:** 473.
35. **Marshall, W. H., & R. W. Gerard**
1933. J. Physiol. **104:** 586.
36. **Hursh, J. B.**
1939. J. Physiol. **127:** 131.
37. **Gasser, H. S.**
1937. In: Gasser & Erlanger, Electrical Signs of Nervous Activity. Univ. of Pennsylvania Press. Philadelphia.
38. **Gerard, R. W.**
1927. J. Physiol. **82:** 381.
39. **Hill, A. V.**
1933. Proc. Roy. Soc. London B **113:** 345.
40. **Gerard, R. W.**
1930. J. Physiol. **93:** 337.
41. **Gerard, R. W., & W. H. Marshall**
1933. J. Physiol. **104:** 535.
42. **Kuffler, S. W.**
1946. Ann. N. Y. Acad. Sci. **47** (5). (In Press.)

43. **Kharasch, M. R., R. R. Legault, A. B. Wilder, & R. W. Gerard**
1936. J. Biol. Chem. **113**: 537, 557.
44. **Colowick, S. P., & E. W. Sutherland**
1942. J. Biol. Chem. **144**: 423.
45. **Warburg, O., & W. Christian**
1942. Biochem. Z. **310**: 384.
46. **Boyer, P. D., H. A. Lardy, & P. H. Phillips**
1942. J. Biol. Chem. **146**: 675.
47. **Bailey, K.**
1942. Biochem. J. **36**: 121.
48. **Engelhardt, W. A., & M. N. Lyubimova**
1942. Biochimia. **7**: 205.
49. **DuBois, K. P., H. G. Albaum, & V. R. Potter**
1943. J. Biol. Chem. **147**: 699.
50. **Greville, G. D., & H. Schwann**
1943. Nature. **152**: 81.
51. **Utter, M. F., & C. H. Werkman**
1942. J. Biol. Chem. **146**: 289.
52. **DuBois, K. P., & V. R. Potter**
1943. J. Biol. Chem. **148**: 451.
53. **DuBois, K. P., & V. R. Potter**
1943. J. Biol. Chem. **150**: 185.
54. **Lipmann, F., & L. J. Tuttle**
1944. J. Biol. Chem. **153**: 571.
55. **Feldberg, W., & T. Mann**
1945. J. Physiol. **104**: 8.
56. **Lardy, H. A., & C. A. Elvehjem**
1945. Ann. Rev. Biochem. **14**: 1.
57. **Evans, E. A.**
1944. Ann. Rev. Biochem. **13**: 187.
58. **Szent-Györgyi, A.**
1945. Acta Physiol. Scand. **9**: Suppl. 25.
59. **Hegenauer, A. H., W. O. Fenn, & D. M. Cobb**
1934. J. Cell. Comp. Physiol. **4**: 505.
60. **Tipton, S. R.**
1936. J. Cell. Comp. Physiol. **7**: 433.
61. **Gerard, R. W., & N. Tupikova**
1939. J. Cell. Comp. Physiol. **13**: 1.
62. **Boyer, P. D., H. A. Lardy, & P. H. Phillips**
1943. J. Biol. Chem. **149**: 529.
63. **Schallek, W.**
1945. J. Cell. Comp. Physiol. **26**: 15.
64. **Montigel, C.**
1943. Helv. Physiol. Pharm. Acta. **1**: c. 47.
65. **Dubner, H. N., & R. W. Gerard**
1939. J. Neurophysiol. **2**: 142.
66. **Fenn, W. O.**
1940. Physiol. Rev. **20**: 377.
67. **von Muralt, A.**
1942. Pflüg. Arch. **245**: 604.
68. **Lissak, K.**
1939. J. Physiol. **127**: 263.
69. **Nachmansohn, D.**
1945. Vitamins & Hormones. **3**: 337.
70. **Pearce, J., & R. W. Gerard**
1942. J. Physiol. **136**: 49.

71. **Feldberg, W.**
1945. Physiol. Rev. **25**: 596.
72. **Brown, G. L., & W. Feldberg**
1936. J. Physiol. **88**: 265.
73. **Mackintosh, F. C.**
1938. J. Physiol. **94**: 155.
74. **Brown, G. L., H. H. Dale, & W. Feldberg**
1936. J. Physiol. **87**: 394.
75. **Gerard, R. W.**
1937. Proc. Assoc. Res. Nerv. & Ment. Dis. **18**: 316.
76. **Holmes, E. G., R. W. Gerard, & E. I. Solomon**
1930. J. Physiol. **93**: 342.
76a. **Tobias, J. M., D. B. Clark, & R. W. Gerard**
1942. Fed. Proc. **1**: 85.
77. **Anfinsen, C. B.**
1944. J. Biol. Chem. **152**: 267, 285.
78. **Lipton, M. A.**
Unpublished data.
79. **Cantoni, G. L., & O. Loewi**
1944. J. Pharm. Exp. Therap. **81**: 67.
80. **Mendel, B., & R. D. Hawkins**
1943. J. Neurophysiol. **6**: 431.
81. **Nachmansohn, D., & H. Schnermann**
1945. J. Biol. Chem. **159**: 239.
82. **Kuffler, S. W.**
1945. J. Neurophysiol. **8**: 113.
83. **Hollinshead, W. B., & C. H. Sawyer**
1945. Anat. Rec. **91**: 282.
84. **Armstrong, P. B.**
1943. J. Cell. Comp. Physiol. **22**: 1.
85. **Bülbring, E., & D. Whitteridge**
1941. J. Physiol. **99**: 201.
86. **Buchtal, F., & G. Kahlson**
1944. Acta Physiol. Scand. **8**: 317, 325.
87. **Ashby, W.**
1944. J. Biol. Chem. **156**: 323, 331.
88. **von Muralt, A.**
1945. Experientia. **1**(5): 1.
89. **Torda, C., H. Wolff**
1944. Proc. Soc. Exp. Biol. & Med. **55**: 88.
90. **Liechti, A., A. von Muralt, & M. Reinert**
1943. Helv. Physiol. Acta. **1**: 79.
91. **Wyss, F.**
1944. Helv. Physiol. Acta. **2**: 121.
92. **Nachmansohn, D., & H. B. Steinbach**
1942. J. Neurophysiol. **5**: 109.
93. **Buchtal, F., & J. Lindhard**
1937. J. Physiol. **90**: 82.
94. **Katz, B., & S. W. Kuffler**
Personal Communication.
95. **Marrazzi, A. S., & R. Lorente de Nó**
1944. J. Neurophysiol. **7**: 83.
96. **Rosenblueth, A.**
1944. J. Physiol. **141**: 196.
97. **Skoglund, C. R.**
1945. J. Neurophysiol. **8**: 365, 377.

98. **Gerard, R. W.**
1941. Ohio J. Sci. **41:** 160.
99. **Tobias, J. M., M. A. Lipton, & A. Lepinat**
1946. Proc. Soc. Exp. Biol. & Med. **61:** 51.
100. **Brock, N., H. Druckrey, & W. Loch**
1943. Biochem. Z. **313:** 300.
101. **Shaffer, M., T. H. Chang, & R. W. Gerard**
1935. J. Physiol. **111:** 697.
102. **Abdon, N. O., & N. A. Nielsen**
1937. Skand. Arch. Physiol. **77:** 13.
103. **Patterson, J. M., N. B. Keevil, & E. W. McHenry**
1944. J. Biol. Chem. **153:** 489.
104. **Welch, J. H., & J. E. Hyde**
1944. J. Neurophysiol. **7:** 41.
105. **Sanz, M. C.**
1943. Pflüg. Arch. **246:** 596.

CONCLUSION

Remarks Made at a Dinner in the Hotel Astor, New York, after the Conference on The Mechanism of Nerve Activity, Sponsored by The New York Academy of Sciences, February 9, 1946.

By J. F. Fulton

Sterling Professor of Physiology, Yale University School of Medicine, New Haven, Connecticut

No set speeches have been planned for this evening, since those who arranged the Conference wished to keep our proceedings entirely informal. The Committee has asked me to express our most sincere thanks to The New York Academy of Sciences and, particularly, to Mrs. Miner and her gracious assistants, who have done so much, both before and during the Conference, to make it outstandingly successful.

In this connection, I must also mention the man—and I do this without instructions from the Committee—who originally conceived the idea of having the Conference and who, with Tracy Putnam's energetic backing, has been so largely responsible for working out the details. David Nachmansohn came to the United States, in the summer of 1939, under the sponsorship of the Dazian Foundation, and since this is something of a family party, I will, perhaps, be forgiven for telling you a family secret. The Dazian Foundation had wished to sponsor a physiologist from Europe. David Nachmansohn was chosen, and I can only say that American Physiology has been vastly stimulated by his presence in this country. He and his wife have made a solid place for themselves here; and, in the language of George Eliot, David, "through his mild persistence, has urged Man's thoughts to vaster issues."

We feel particularly fortunate in being able to welcome so many distinguished colleagues from abroad, this having been made possible by the vision of the Rockefeller Foundation and of the Commission for Relief in Belgium. Our colleagues from France bring us heartening news of the revival of their laboratories and of their faith in the universal fellowship of scientific men. We are also happy to see Professor Augusto Pi-Suñer of Barcelona and Caracas, and his son, Dr. Jaime Pi-Suñer. Also, Arturo Rosenblueth from Mexico. In Doctors Höber and Michealis, we have distinguished representatives of the

highest traditions of German science, men who could never surrender their faith in academic freedom.

Frédéric Bremer exemplifies all that we most admire in his countrymen: loyalty, self-reliance, humor, industry, and, with it all, a burning zeal for research that sustained him in his vigorous way of life during the lean years through which he has just passed. When conditions in his university laboratory made work impossible in 1943, he retired to the cellar of his house for nearly two years, and there continued his experimentation and his writing. Virtue cannot be enhanced by calling it to public notice and I do so now, not to add luster to Frédéric Bremer's position in the world of science, but rather that others may take inspiration from his faith and his example. He states modestly that he has merely followed in the pathways of those he seeks to emulate. Many of us here share with him a common devotion to two of his masters: one was Harvey Cushing; the other, Sir Charles Sherrington. Sir Charles, whom Bremer has recently seen, is now in his ninth decade. Bremer found him hard at work, bringing out a biography of Jean Fernel, the sixteenth century physician and humanist; while he is somewhat crippled by arthritis, his mind remains ever vigorous.

Another pupil of Sherrington is John Eccles, whose industry, like that of Bremer's, is phenomenal. In a very short space of time, he has had eight children, and, not content with bringing up a large family at home, he also created a laboratory family of loyal associates: Hebbel Hoff and David Lloyd who were his pupils at Oxford, and Stephen Kuffler (whom you have all enjoyed hearing at the Conference), Bernard Katz, and many younger men whose names we are beginning to see in the literature. Chandler Brooks of Baltimore permits me to tell you that he, too, is going presently to New Zealand, to experience for a year the stimulating atmosphere of Eccles' laboratory. Characteristic of the Eccles family, Mrs. Eccles has extended a cordial invitation to Dr. and Mrs. Brooks to live with them while they are in Dunedin, should they have difficulty in finding accommodations. I could tell you more about Jack Eccles, but since he is a good friend of mine I shall spare him, the more so since he knows much too much about me to make it safe to indulge in blackmail.

My pleasant duty in closing is to propose a toast. Since we did not wish to obligate anyone to speak, it seemed inappropriate to single out our guests, for they have been one with the Conference. But it has seemed highly appropriate to ask you to drink a standing toast to the man who has probably influenced our thinking more profoundly than anyone now living—Sir Charles Sherrington.

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
VOLUME XLVII, ART. 5. PAGES 603-664
MAY 9, 1947

PHYSIOLOGICAL AND PSYCHOLOGICAL FACTORS IN SEX BEHAVIOR*

BY

S. BERNARD WORTIS, GREGORY BATESON, WILLIAM E. GALT, MORRIS HERMAN, ALFRED C. KINSEY, AND WILLIAM C. YOUNG

CONTENTS

PAGE

INTRODUCTION TO THE CONFERENCE. BY S. BERNARD WORTIS 605

PART I: ANIMAL SEX BEHAVIOR

ANIMAL ENDOCRINES IN RELATION TO SEXUAL BEHAVIOR. BY WILLIAM C. YOUNG 609

SEX BEHAVIOR IN PRIMATES. BY WILLIAM E. GALT 617

PART II: HUMAN SEX BEHAVIOR

SEX BEHAVIOR IN THE HUMAN ANIMAL. BY ALFRED C. KINSEY 635

ABERRANT SEX BEHAVIOR IN HUMANS. BY MORRIS HERMAN 639

SEX AND CULTURE. BY GREGORY BATESON 647

* This series of papers is the result of a Conference on Physiological and Psychological Factors in Sex Behavior held by the Sections of Biology and Psychology of The New York Academy of Sciences, March 1, 1946.

COPYRIGHT 1947

BY

THE NEW YORK ACADEMY OF SCIENCES

INTRODUCTION TO THE CONFERENCE

By S. Bernard Wortis

Professor of Psychiatry, New York University College of Medicine, and Director, Psychiatric Division, Bellevue Hospital, New York, N. Y.

Attitudes and traditions in our society, with regard to sex behavior, are derived from needs and ideas that have served man at some time in his development. Frequently, some of these attitudes and traditions become codified in law and, subsequently, a frozen segment of our mores. This conference will bring forth new material and point to correlations that emphasize the need to change our attitudes regarding human sexual behavior. One hopes that it may help to modify our laws in these matters.

Sexual function in man has come to be dominated by his mind. In lower animals, the endocrine secretions play a most important role in sexual behavior. In infra-human primates, the endocrine factor is less controlling. Sexual behavior of the monkey and the chimpanzee is modified by factors concerned with personal dominance, aggressiveness, and the individual's status in his social group. In man, these social, cultural, and personal psychodynamic factors have become preeminent. This "encephalization" of sexual function has developed, in evolutionary time, to the point where some endocrine secretions are of secondary importance to the mind. Today, we recognize that psychologic-cultural conditioning factors play a vital role; that sexual attitudes are developed in early life (before the child acquires specific factual knowledge) and are important in the psychosexual adjustment of the adolescent and the adult. We must, however, modify our ideas concerning normal patterns of sexual behavior.

Biologists tell us that neuromuscular patterns of bisexuality are present in the nervous systems of primates and other mammals. Even the endocrine patterns of humans show "male" and "female" sex hormones to be present in every person. Moreover, variations in sexual behavior, such as mouth-genital contact and sexual contact between primates of the same sex, have been observed in animal studies. Social taboos have strongly colored the attitude of our society to such experiences in man, despite the fact that such activity is usual in primate mammalian sex behavior. It is time that our society modified its attitudes to variations of the sexual impulse and tempered its laws relating to these matters with biological and psychodynamic wisdom. Much of the evidence collected in these papers will provide the basis for such orientation.

PHYSIOLOGICAL AND PSYCHOLOGICAL FACTORS IN SEX BEHAVIOR

PART I

ANIMAL SEX BEHAVIOR

ANIMAL ENDOCRINES IN RELATION TO SEXUAL BEHAVIOR

By William C. Young

*Cedar Crest College, Allentown, Pennsylvania**

What is unknown about the endocrines and sexual behavior in mammals still exceeds by far what is known. It was only 10 years ago that the view of most endocrinologists was confined to the generalizations that the manifestation of male sexual behavior is stimulated by androgens, and that female sexual behavior is stimulated by estrogens. Much has been learned during the decade since 1936, but quite as important for the future, it seems to me, is the acquisition of a sense of direction. It is this theme that is to be developed in what follows.

The relationship of the endocrines to sexual behavior has been studied, in males and females, in a variety of mammals: rats, mice, guinea pigs, and hamsters, among the rodents; rabbits, cats, sheep, horses, cows, and dogs, among the larger domesticated mammals; and the *Rhesus* monkey, bonnet macaque, and chimpanzee, among the laboratory primates.

As far as the male is concerned, but few complications have thus far arisen. Agreement is general, among those whose work Beach[3] has reviewed, that the sexual responsiveness of castrate individuals is restored by the administration of androgenic substances; that the sexual excitability of sluggish individuals is increased by similar treatment; and that prepuberal rats likewise respond to androgenic action.

In the female, the situation is more complex. The results from experiments on the large domesticated, and the larger laboratory mammals, which are reviewed elsewhere,[36] have led to the opinion that mating behavior in the female is normally stimulated by such naturally occurring estrogens as estradiol or estrone, although in some members of this group, notably the sheep, a complete physiological response has not been obtained regularly, and the conclusion was reached by Cole, Hart, and Miller[11] that either a proper balance of hormones was not attained or unknown factors were involved.† In laboratory

* Present address: Department of Anatomy, University of Kansas, Lawrence, Kansas.

† It has recently been reported, however (**Fraps, Schott, Simmons, & Phillips.** Abstract, Proc. Am. Soc. Zool., Anat. Rec. **98**: 74. 1946), that, in the sheep and goat, progesterone has the effect of terminating estrous periods induced by estradiol benzoate, and that the inconstant results obtained by other investigators may, perhaps, be explained by the presence of functional *corpora lutea* which developed following the injections of gonadotrophins that had been made in attempts to induce estrus.

rodents such as the rat, mouse, guinea pig, and hamster, mating responses similar to those shown prior to ovariectomy have been experimentally produced only when a priming or conditioning injection of an estrogen was followed by the injection of a small amount of progesterone.[12 7 39, 2, 27, 16]

An additional and previously unreported fact was revealed while the experiments which led to this latter conclusion were in progress. The sexual behavior pattern, as seen in the guinea pig and rat, may have as many as three components: the mating response, a male-like mounting activity, and a certain amount of running activity. It is known for the mating response in the spayed rat, and for the mating response and male-like mounting activity in the spayed guinea pig, that a single treatment with estradiol benzoate, or estrone followed by progesterone, is sufficient to induce these behaviors in an amount comparable with that shown prior to ovariectomy. A normal amount of running activity, on the other hand, has not been produced by this treatment. This has been accomplished only when small quantities of estrone were taken in the drinking water,[25] or administered daily in the form of injections of the aqueous hormone,[26] or when pellets of this hormone were implanted subcutaneously.[40] In the case of this component of the sexual behavior pattern in the rat, estradiol benzoate was but slightly effective (a single injection was entirely inadequate), and progesterone appears to have been without any action.

To what extent can other endocrine substances or synthetic hormones be substituted for the naturally occurring hormones? As far as I have been able to ascertain, experiments in which efforts have been made to find substitutes for estrone in the induction of spontaneous running activity, have been performed only by Hemmingsen and Krarup.[19] The synthetic hormone they used (9:10-dihydroxy-9:10-di-n-propyl-9:10-dihydro-1:2:5:6 dibenzanthracene) raised the level of activity well above that shown by the ovariectomized rats they injected, but there is nothing to indicate whether or not the level shown prior to ovariectomy was restored.

In the case of the mating response, stilbestrol[13] and triphenyl ethylene[28] can replace the naturally occurring estrogens; testosterone propionate can, to some extent, substitute for estrogen in rats[3]; and desoxycorticosterone acetate,[32] pregneninolone,[31] or an aqueous extract of the adrenal cortex can take the place of progesterone. However, in every report in which the point is covered, the naturally occurring hormones seem to have been more effective.

At least two investigators have shown that the frequency with which male-like mounting activity is displayed is increased by treatment with androgens,[1, 3] but in experiments which involved the induction of mounting activity by a single injection of gonadal hormones, the androgens used by Young and Rundlett[39] did not have a synergistic action comparable to that of progesterone.

The conclusion to be drawn from these experiments is that there are substances which come close to substituting for the naturally occurring estrogens and progesterone, in the production of the mating pattern characteristic of the individual animal. That the substances which have been used may substitute completely for the naturally occurring ovarian hormones, is a question which I feel is still unanswered.

Thus far, it has been shown that certain gonadal hormones are active in the production of sexual behavior. In moving beyond this point, I will elaborate somewhat on an aspect of the subject which in the aggregate has received considerable attention, but which nevertheless is in need of further, and a more systematically organized, study. Although the rapidly accumulating literature has never been reviewed, it is common knowledge that members of a species vary greatly in the nature and extent of their responsiveness, sensitivity, or reactivity to hormonal substances. This is as true of the behavioral responses as it is of the tissue responses. As the sexual behavior of the laboratory and large domesticated mammals has been studied, very great differences have been found in the qualitative and quantitative aspects of their behavior. Many of us have sought to correlate these differences, to ascertain if an animal in which the mating reactions were especially intense was also active in mounting other animals,[37] or if a prolonged period during which the female would accept the male was associated with a great amount of running activity.[18, 19] It cannot be said that such correlations have been found. Efforts have been made to ascertain if any relationship exists between the amount of estrous behavior and the number of maturing follicles[38, 40] and, in the case of two chimpanzees displaying contrasting extremes of behavior, the amount of excreted estrogenic substances was assayed.[15] However, in neither set of observations was any correlation shown.

Tentatively, the conclusions have been reached for the lower mammals[40] that "the mating response, the male-like mounting activity and the running activity are stimulated by the cyclically produced secretions of the ovaries rather than by a single substance. But in addition to the factor of a multiplicity of hormones having a selectivity or specificity of action, the situation is complicated by

the circumstance that there is not a common threshold or sensitivity for the three parts of the pattern. Many observations suggest that each of the mechanisms mediating the three parts of the total pattern responds in its own manner and that the behavior which an animal displays is more strongly influenced by the sensitivity of these mechanisms to hormonal substances than by any quantitative fluctuations in the amount of secreted substances, provided the threshold of stimulation has been reached."

This inclination to focus attention on the sensitivity or responsiveness of the tissues on which the endocrines act as a central problem, impels us to seek factors which are responsible for the differences in sensitivity. This, it would seem, is one of the directions in which we must move if the basis of qualitative and quantitative differences in sexual behavior patterns is to be found.

The possibility that genetic factors may have much to do with the character of the response is the first to suggest itself. Although evidence for the existence of genetic factors within a species is still not extensive,[29, 10] no one would question the suggestion that the behavior patterns displayed by two species such as the rabbit and cat have a genetic basis. Proof is lacking, but it probably is no less reasonable to suppose that genetic factors operate within a species to produce strain differences, such as those described by Hauschildt and Evans.[17]

There is ample evidence for the existence of a relationship between the age of the animal and its responsiveness to hormonal substances, which comes from the work of those who have studied tissue reactions and those who have studied behavior,[20, 22, 23, 24, 30] and from older articles cited by Young.[36] For the most part, though, in the laboratory rodents, the changes which are in the direction of an increasing sensitivity seem to be confined to the first month or two of postnatal life. Thereafter, through the tenth month in the rat, and the second year in the guinea pig, no further changes have been detected.[35] It remains to be shown, therefore, for the period when reproductive activities are possible, and later, that age is related to the sensitivity to gonadal hormones and thereby to the character of estrous behavior.

Seasonal fluctuations in sensitivity of the tissues to gonadal and hypophyseal hormones are asserted to occur,[6, 9, 14, 21] but there is no very satisfactory evidence that the sensitivity to hormonal stimulation, involving behavioral responses, varies similarly.

The relationship of the endocrines to the problem of sensitivity needs clarification and further study. Hypophysectomy has, in general, been followed by difficulty in obtaining estrous responses,

although hypophysectomized rabbits, dogs, cats, and rats have been brought into heat by replacement therapies. The high percentage of failures, and the weakness of the responses when estrous behavior was obtained, could be attributed to the severe general effects of the operation. The procedure, moreover, does not enlighten us with respect to the effects of specific endocrine organs such as the adrenal gland, the thyroid, and the gonads themselves. As far as the adrenals are concerned, the most suggestive studies are those of Bourne and Zuckerman,[8] in which evidence is cited for the action of the adrenal cortex in lowering the threshold for estrogen stimulation, as measured by changes within the vaginal epithelium. Tissue sensitivity has also been said to be modified in experimental hyper- and hypotension.*

Of the investigations dealing with the relationship of the endocrines to the sensitivity or responsiveness of the elements mediating behavioral changes, studies of the gonads have been the most numerous. What is clearest is that, during the prenatal period and much of postnatal life, the gonads are without any detectable effect. Not only are the development and maintenance of a normal sensitivity to estrogen independent of ovarian action,[35] but the injection of large quantities of estrogens and androgens during early postnatal life actually prevents the acquisition of a responsiveness to the estrus-inducing hormones.[33, 34] In a confirmation and extension of this conclusion, Beach[5] was able to show that the differentiation and development of the structures controlling sexual behavior are possible, even under the conditions of a congenital absence of gonadal tissue. Such being the case, we are compelled to look elsewhere than to the gonads for the factor which is responsible for this change in sensitivity. Moore[22] has postulated that sexual differentiation is under the control of genetic sex-determining factors, rather than of gonad-secreted hormones and, without denying a later activation to such hormones,

* Two recent abstracts contain information bearing on this point. **Fleischmann** (Fed. Proc. 5 (2): 28. 1946) notes that the ability of estrogen to modify certain metabolic changes in immature chicks is neutralized by treatment with thyroxin, whereas estrogenic effects on the structure of the oviducts are unaffected. In an investigation (**Young, Webster, & Innes.** Abstract, Proc. Am. Soc. Zool., Anat. Rec. January, 1947; in press) in which the sensitivity of the behavioral reactions was tested under the conditions of acute hyperthyroidism, the responses given by ovariectomized guinea pigs which survived the treatment were quantitatively and qualitatively similar to those shown prior to the injection of crystalline thyroxin. In three older studies (**Reiss & Perény.** Endokrinologie **2:** 181. 1928. **Van Horn.** Endocrinology **17:** 152. 1933. **Meyer & Wertz.** Proc. Soc. Exp. Biol. & Med. **38:** 843. 1938), the responsiveness of the vaginal epithelium to estrogens was shown to be lower in ovariectomized rats to which thyroid preparations had been given. In a third recent study, **Langham & Gustavson** (Fed. Proc. 5 (2): 143. 1946) found that injections of dl-thyroxin into ovariectomized rats for periods up to 10 days greatly decreased the sensitivity of the vaginal epithelium to constant doses of estrone, and that thyroparathyroidectomy and thiourea treatment had the ultimate effect of increasing sensitivity.

it is suggested that these unidentified non-gonadal factors may account for many of the differences between individuals which are encountered everywhere. The possibility that such an inherited sensitivity can be modified, or influenced by physiological as well as by environmental factors, must, of course, be studied even more carefully than has been done. However, with excellent end-points available in both males and females, this should not only be practicable, but should also prove a fruitful field of inquiry.

LITERATURE CITED

1. **Ball, J.**
 1940. The effect of testosterone on the sex behavior of female rats. J. Comp. Psychol. **29**: 151–165.
2. **Beach, F. A.**
 1942a. Importance of progesterone to induction of sexual receptivity in spayed female rats. Proc. Soc. Exp. Biol. & Med. **51**: 369–371.
3. **Beach, F. A.**
 1942b. Male and female mating behavior in prepuberally castrated female rats treated with androgens. Endocrinology **31**: 673–678.
4. **Beach, F. A.**
 1942c. Analysis of factors involved in the arousal, maintenance, and manifestation of sexual excitement in male animals. Psychosomatic Med. **4**: 173–198.
5. **Beach, F. A.**
 1945. Hormonal induction of mating responses in a rat with congenital absence of gonadal tissue. Anat. Rec. **92**: 289–292.
6. **Belcher, J. C.**
 1940. Seasonal and Experimentally Induced Changes in the Reproductive Tract of the Female Bat *Myotis grisescens*. Thesis. University of Missouri.
7. **Boling, J. L., & R. J. Blandau**
 1939. The estrogen-progesterone induction of mating responses in the spayed female rat. Endocrinology **25**: 359–364.
8. **Bourne, G., & S. Zuckerman**
 1941. The influence of the adrenals on cyclical changes in the accessory reproductive organs of female rats. J. Endocrinology **2**: 268–282.
9. **Bradbury, J. T.**
 1944. The estrous rabbit as a quantitative assay animal. Endocrinology **35**: 317–324.
10. **Brody, E. G.**
 1942. Genetic basis of spontaneous activity in the albino rat. Comp. Psychol. Monog. **17**(5).
11. **Cole, H. H., G. H. Hart, & R. F. Miller**
 1945. Studies on the hormonal control of estrous phenomena in the anestrous ewe. Endocrinology **36**: 370–380.
12. **Dempsey, E. W., R. Hertz, & W. C. Young**
 1936. The experimental induction of oestrus (sexual receptivity) in the normal and ovariectomized guinea pig. Am. J. Physiol. **116**: 201–209.
13. **Dodds, E. C., W. Lawson, & R. L. Noble**
 1938. Biological effects of the synthetic oestrogenic substance 4:4′-dihydroxy-α:β-diethylstilbene. Lancet **234**: 1389–1391.

14. **Duszyńska, J.**
1939. Seasonal changes in sensitivity to oestrone. Acta Biol. Exp. (Warsaw) **12**: 229–237.

15. **Fish, W. R., W. C. Young, & R. I. Dorfman**
1941. Excretion of estrogenic and androgenic substances by female and male chimpanzees with known mating behavior records. Endocrinology **28**: 585–592.

16. **Frank, A. H., & R. M. Fraps**
1945. Induction of estrus in the ovariectomized golden hamster. Endocrinology **37**: 357–361.

17. **Hauschildt, J. D., & J. S. Evans**
1942. Difference in response of certain strains of rats to augmentative gonadotropic effect. Proc. Soc. Exp. Biol. & Med. **50**: 353–355.

18. **Hemmingsen, A. M.**
1933. Studies on the oestrus-producing hormone (oestrin). Skand. Arch. Physiol. **65**: 97–250.

19. **Hemmingsen, A. M., & N. B. Krarup**
1937. The production of mating instincts in the rat with chemically well defined oestrogenic compounds. Det. Kgl. Danske Vidensk. Selskab., Biolog. Meddel. **13**(8).

20. **Hooker, C. W.**
1942. Pubertal increase in responsiveness to androgen in the male rat. Endocrinology **30**: 77–84.

21. **McCormack, G., & C. A. Elden**
1945. Seasonal variation of rabbits to pituitary extracts. Endocrinology **37**: 297–299.

22. **Moore, C. R.**
1941. On the role of sex hormones in sex differentiation in the opossum (*Didelphys virginiana*). Physiol. Zool. **14**: 1–45.

23. **Moore, C. R., & C. F. Morgan**
1943. First response of developing opossum gonads to equine gonadotropic treatment. Endocrinology **32**: 17–26.

24. **Price, D., & E. Ortiz**
1944. The relation of age to reactivity in the reproductive system of the rat. Endocrinology **34**: 215-239.

25. **Richter, C. P.**
1934. Pregnancy urine given by mouth to gonadectomized rats: Its effect on spontaneous activity and on the reproductive tract. Am. J. Physiol. **110**: 499–512.

26. **Richter, C. P., & C. G. Hartman**
1934. The effect of injection of amniotin on the spontaneous activity of gonadectomized rats. Am. J. Physiol. **108**: 136–143

27. **Ring, J. R.**
1944. The estrogen-progesterone induction of sexual receptivity in the spayed female mouse. Endocrinology **34**: 269–275.

28. **Robson, J. M., & A. Schönberg**
1937. Oestrous reactions, including mating, produced by triphenyl ethylene. Nature **140**: 196.

29. **Rundquist, E. A.**
1933. Inheritance of spontaneous activity in rats. J. Comp. Psychol. **16**: 415–438.

30. **Selye, H., & S. Albert**
1942. Age factor in responsiveness of pituitary and adrenals to folliculoids. Proc. Soc. Exp. Biol. & Med. **50**: 159–161.

31. **Soderwall, A. L.**
1940. Induction of sexual receptivity in estrogen-conditioned spayed female guinea pigs by orally administered progesterone and pregneninolone. Endocrinology **27**: 840–841.

32. **van Heuverswyn, J., V. J. Collins, W. L. Williams, & W. U. Gardner**
1939. The progesterone-like activity of desoxycorticosterone. Proc. Soc. Exp. Biol. & Med. **41**: 552–554.

33. **Wilson, J. G.**
1943. Reproductive capacity of adult female rats treated prepuberally with estrogenic hormone. Anat. Rec. **86**: 341–363.

34. **Wilson, J. G., W. C. Young, & J. B. Hamilton**
1940. A technic suppressing development of reproductive function and sensitivity to estrogen in the female rat. Yale J. Biol. & Med. **13**: 189–202.

35. **Wilson, J. G., & W. C. Young**
1941. Sensitivity to estrogen studied by means of experimentally induced mating responses in the female guinea pig and rat. Endocrinology **29**: 779–783.

36. **Young, W. C.**
1941. Observations and experiments on mating behavior in female mammals. Quart. Rev. Biol. **16**: 135-156, 311-335.

37. **Young, W. C., E. W. Dempsey, & H. I. Myers**
1935. Cyclic reproductive behavior in the female guinea pig. J. Comp. Psychol. **19**: 313–335.

38. **Young, W. C., E. W. Dempsey, H. I. Myers, & C. W. Hagquist**
1938. The ovarian condition and sexual behavior in the female guinea pig. Am. J. Anat. **63**: 457–487.

39. **Young, W. C., & B. Rundlett**
1939. The hormonal induction of homosexual behavior in the spayed female guinea pig. Psychosomatic Med. **1**: 449–460.

40. **Young, W. C., & W. R. Fish**
1945. The ovarian hormones and spontaneous running activity in the female rat. Endocrinology **36**: 181–189.

SEX BEHAVIOR IN PRIMATES

By William E. Galt
The Lifwynn Foundation, Westport, Connecticut

INTRODUCTION

The study of sex behavior in infra-human primates offers a fertile field of investigation. Monkeys and apes are relatively so close to man, from the point of view of evolutionary kinship, structural organization, and functional capacity, that significant findings in regard to the psychobiological interrelations of sex behavior at this level should provide valuable leads and insights in studying human sex behavior and its dysfunctions. Furthermore, the sex behavior of the sub-human primate, unlike that of man, is relatively direct, frank, and free of self-conscious attributes. It can, therefore, be studied under more or less controlled, experimental conditions.

The study of socio-sexual behavior and interrelations in these animal types is, however, not without its inherent difficulties. Several general lines of approach have been followed in studies dealing with this topic, but greatest emphasis in the present review will be given to findings which have resulted from investigations of sex behavior where the animals live under laboratory conditions, and where the answer to a specific question is sought, other variables being controlled as nearly as possible. It should be borne in mind that, in such studies, the animal is necessarily placed in a socially, sexually, and bionomically artificial situation. Conclusions drawn from the results must, therefore, be tempered by these considerations. Certainly, such conditions as captivity, segregation, and the stimulus of an observer intensify the primary sexual activity of infra-human primates.

SEXUAL CYCLE IN INFRA-HUMAN PRIMATES

The sexual cycle in the chimpanzee, and in some types of monkeys, lends itself particularly well to study, since it contains several easily observed phases. In the chimpanzee, the following macroscopic phases are discernible: menstrual bleeding; a postmenstrual condition, which is followed by tumescence of the genital area; a phase of maximal swelling, followed by detumescence; and a premenstrual phase, in which the sexual skin is quiescent. The sexual cycle in the monkey averages 28 days, while the median for the chimpanzee cycle is 35

days.[10] The longer cycle in the chimpanzee is due to a greater length in the pre-swelling and swelling phases. The combined length of these two phases in the chimpanzee is about 25 days, and is more than twice the length of comparable phases in the monkey. Evidence indicates that the swelling and coloration of the sexual skin in monkey and chimpanzee is caused by the female hormone, estrin. Ovariectomy performed on monkeys during the phase of maximal swelling results in prompt subsidence of the sexual skin, while experimental injection of estrin into ovariectomized monkeys produces the characteristic enlargement and coloration.

TIME OF OVULATION

Ovulation occurs in monkeys at the time of maximal genital swelling, about mid-way between two menstrual periods.[1] Elder[18] has demonstrated, using the method of experimental matings, that, in the chimpanzee, pregnancy results only when copulation occurs during the last few days of maximal genital swelling. Presumably, then, ovulation is limited to this period, and is followed rather promptly by detumescence, as is the case in the baboon and in the monkey.[41,42,44] Finch, Yerkes, and Elder,[20] using the Burr-Lane-Nims electrical potential technique, tested a female chimpanzee throughout two sexual cycles and found a maximal potential, presumably indicating ovulation, on the 17th day of each cycle, which was within the last two days of maximal swelling. Additional evidence on time of ovulation in chimpanzees is provided by Young and Yerkes.[40] These investigators examined the ovaries of four adult chimpanzees of known genital status and found that ovulation occurred shortly before detumescence.

SEXUAL RECEPTIVITY

Early literature on the sex behavior of infra-human primates advanced the view that they differed from many other mammals, and were similar to man, in that the male and female may mate at any time during the sexual cycle, instead of only during brief cyclic periods of estrus. Maslow[25] explains this appearance of copulatory behavior throughout the sexual cycle of the monkey as being due to the superimposing of the factor of dominance, which has, in these animals, become closely interrelated with the sexual response. He noted a qualitative difference between copulatory behavior in monkeys, depending upon whether the female was in estrus or in a physiologically

non-receptive phase of her cycle. Copulation during the period of estrus was more apt to be of the functional type, whereas at other times it was more likely to be of the dominance type. Yerkes[32] holds that a basic cyclicity of behavior prevails in the sexual life of infra-human primates, and that this cyclicity is obscured by dominance and other factors of a socio-sexual nature.

There are definite indications, certainly, that sexual behavior in monkeys and apes is powerfully oriented in regard to the period of female estrus and ovulation. Ball and Hartman[1] have shown that, in the monkey, there is an intensification in copulation and other sexual activity immediately preceding ovulation. Zuckerman[42] has likewise observed that male baboons show a preference for females in which the sexual skin is swollen. On the basis of the observation of some six hundred controlled matings, each of which represented a definitely prearranged experiment with two subjects of known history, sexual status, and relationship, Yerkes,[32] and Yerkes and Elder[19 36 37] conclude that the female chimpanzee is non-receptive except during the swelling phases of her sexual cycle. They report that copulation in typical and complete form among mature and congenial animals occurred only during estrus. Sexual desire in the male was also found to be correlated with the sexual status of the female. It is lowest during the menstrual and adjacent cycle phases, and is highest during the phase of maximal swelling. The number of copulations observed during maximal swelling was twice as great as during the remainder of the sexual cycle, although the maximal swelling phase is not more than one-third the duration of the total cycle. Behavioral data indicated that, when the associated female was in other than the maximal swelling phase of her cycle, the male's desire for copulation was weak, and that if copulation occurred it proved to be less adequately satisfying to him. The acceptability of the female to the male thus varies directly with her receptivity.

Varied socio-sexual factors have been found to influence the length of the period of the sexual cycle during which the female infra-human primate will copulate. As has been mentioned earlier, important among these is the factor of dominance. It appears that the more dominant the male, both individually and as a phylogenetic characteristic of the animal type, the wider is the range of copulatory behavior in the typical sexual cycle, and the greater is the tendency of the female to respond accommodatingly to the advances of the male, irrespective of her sexual status. Some of the other factors which have been found to be effective in bringing about copulation in

chimpanzees at other times than when the female is typically receptive, are: prior isolation of the male from the female; lack of opportunity for the male to observe the sexual status of the female before being placed with her; immaturity, or lack of sexual experience, of male or female; generalized environmental excitement; or the presence of a human observer. Notwithstanding that the intrusion of such socio-sexual factors clouds the issue, Yerkes, on the basis of his wealth of observation of chimpanzee mating behavior, makes the following generalization: "Normally and typically, a multiparous female will mate with a male with whom she is intimately acquainted and on friendly or affectionate terms only during a few days of the period of genital swelling about and including the usual time of ovulation. At other times, the female as a rule is ignored sexually by the male, or, if approached and solicited, she avoids him."

COPULATORY BEHAVIOR

Behavior exhibited in the act of copulation is exceedingly variable in the infra-human primates. It not only varies from one species to another, but from animal to animal in the same species, and in the same animal during different phases of the sexual cycle. The dorso-ventral copulatory position, with the male covering the female, is the one typically assumed in these primates. Ventro-ventral mounting has been reported, however, to occur on occasion, both in monkeys and in the great apes. Yerkes has emphasized the qualitative differences in the copulatory act of the chimpanzee, depending upon whether it takes place during or outside the period of female estrus. If the female is fully receptive and sexually experienced, she usually takes the initiative and goes to the male when their access to one another is permitted. The male typically assumes copulatory posture and awaits the female's approach and presentation. Upon reaching him, the female crouches low or even flat upon the ground, with limbs flexed and genitals directed toward, and often pushed against, the male. Completion of copulation is often followed by mutual examination of genitalia, and by grooming.

During times when the female is not fully receptive, the male chimpanzee usually goes to the female, on their having access to one another. She may respond by presenting, but the presentation is high and differs markedly from the intermediate or low presentation characteristic of the fully receptive female in maximal swelling. It is usually not followed by copulation. Although the male is physically

capable of forcing his sexual desires upon the female, regardless of her desire and sexual status, such behavior has been found to occur infrequently in chimpanzees if the mates are known to each other, are congenial, and if the male has had an opportunity to observe the sexual status of the female before being placed with her.

There are indications in the sex life of both monkeys and chimpanzees that the female has greater copulatory power than has the male. In his observation of free-ranging howling monkeys, Carpenter[12] found that, when a mature male was with a receptive female, copulation occurred at frequent intervals until the male was satiated, when, in many cases, he was promptly replaced by another. As many as three males were observed to be satiated in succession by one estrus female. Evidence of sexual satiation in the female was, however, not observed. Yerkes[35] contributes similar observations on the chimpanzee. He finds that, during estrus, the capacity of the female for copulation is almost unlimited, whereas the male is readily satiated, fatigued, or wholly exhausted by one female.

Males differ in their acceptability to the female as sexual partners. Carpenter[12] reports that the female howling monkey may refuse a given male and immediately accept another. Yerkes[32] found that such preferential behavior also occurred in chimpanzees, and that with a preferred sexual partner the period of receptivity of the female tended to lengthen. Factors in the male chimpanzee which appear to be positively correlated with his sexual acceptability are vigor, penis length, copulation time, and number of pelvic thrusts before orgasm. A high degree of sexual selectiveness, aggressiveness, or dominance on the part of the male made him less acceptable to the typical female studied. It is of interest, in this latter connection, that, in a recent study of factors responsible for a satisfactory sexual *rapport* between man and woman, Maslow[27] found that marriages were more satisfactory where the male and female were of comparable dominance-drive, or where the male was slightly more dominant. Successful marriages did not usually issue from the union of a very dominant and a very submissive individual.

RELATIONSHIP OF DOMINANCE TO SEX

It was early noted, in the observation of sexual behavior in monkeys, that characteristic responses of a sexual nature subserved functions other than that of reproduction. Studies by Hamilton[22] and Kempf,[23] in which sexually mature *Macacus* monkeys and baboons were placed

together, showed that the female might assume the position of sexual presentation seemingly in an effort to appease an angry, aggressive male. Tinklepaugh[31] reports that sexual responses were frequently elicited, in the monkeys and chimpanzees which he studied, in situations where thwarting occurred, and that these responses seemed to be of a substitutive nature. Furthermore, such substitutive responses are not limited to the female of the species. A subordinate male monkey may also take the position characteristic of the presenting female in response to a dominant, aggressive male. Such behavior has been described by Zuckerman[42] in baboons, and was observed many times in *Macacus rhesus* and Sooty Mangabey monkeys, by Warden and myself, in experiments securing data on dominance behavior in these animals. Zuckerman regards sexual presenting of this type as an important means by which a subordinate animal adjusts itself to a system based on dominance.

In a series of studies, Maslow,[24-26] and Maslow and Flanzbaum[28] made a systematic investigation of the effect of dominance on the social and sexual behavior of various types of monkeys. It was found that, while presentation does occur more frequently in a subordinate than in a dominant animal, it is by no means the most constant characteristic of subordinate behavior. In fact, the subordinate animal may often completely fail to show this reaction. A sexual aspect observed to be much more highly correlated with dominance status was dorso-ventral mounting of one animal by another. Maslow found that the dominant animal in a pair was responsible for between 90 and 98 per cent of the mountings, irrespective of the sexual gender of either animal. As a result of his studies, Maslow concluded that dominance is an extremely important determinant of social and sexual behavior in the monkey.

Maslow made some incidental observations, which indicated that a female naturally dominant to the male with which she was paired would, during the period of estrus, become subordinate to the male. Not only did the male mount her at such times, but he also tended to assume dominance in other characteristics of their relationship. Zuckerman[42] has reported this same type of recession in dominance of a dominant female during estrus.

Yerkes and his associates of the Yale Laboratories of Primate Biology[17,30,33,34] have contributed the most extensive experimental work on the problem of the shifting of dominance with changes in the female sexual cycle. In these studies, attention was concentrated on forms and aspects of behavior, and especially on sexual responses, as

elicited by an experimental food-getting situation. The technique was essentially the same in the several studies made. Two experimental animals were confined in a large cage and secured food either by way of a chute or by opening a box on a platform in front of the cage. The pairs of animals were usually caged together throughout a series of observations which continued for at least one complete sexual cycle of the female. Observations were made daily during the spaced presentation of ten pieces of preferred food. Since only one of the two subjects was able to secure food on a given trial, the situation afforded excellent opportunity for the exhibition of various degrees of dominance and subordination behavior.

The results of the several studies may be briefly summarized as follows: In the male-female pairings, the priority of response in the food-getting situation typically shifted from the dominant male to the subordinate female during the female's period of estrus or maximal genital swelling. With the onset of detumescence in the female, priority of response was again assumed by the male, and he controlled food-taking until the female again reached estrus. From his extensive experience in the observation of chimpanzee behavior, Yerkes describes the priority of response which the female exhibits in regard to food, when sexually receptive, as a "right" or "custom" of chimpanzee interchange which is tacitly accepted by both members of the pair. The interpretation is advanced that, in general, the assumption of priority of response on the part of the female, during the period of sexual receptivity, is in exchange for sexual accommodation of the male. This, however, may not be the effective causation in the specific case, as is shown in the results obtained for one female who, although habitually refusing to mate even during periods of maximal receptivity, nevertheless gained priority of response at such times.

In the female-female pairings, changes in response priority usually occurred during the various phases of the menstrual cycle, but the relationships were by no means as clear as with the male-female pairs. Sexual receptivity appeared to have an opposite effect upon behavior, depending upon whether it was the naturally dominant or the naturally subordinate member of the pair that was in this physiological state. The naturally dominant female tended, in estrus, to lose priority of response to the subordinate female, while the naturally subordinate female, when sexually receptive, ordinarily achieved priority of response over the dominant female. This latter relationship was shown more clearly in a later study in which ovariectomized

females were paired with other females in order to eliminate cyclic sexual changes in one of the members of each pair.

An attempt to determine the physiological basis for the observed shifts in priority of response with sexual status has been made by Clark and Birch.[16] These investigators studied the effects of differential sex-hormone level upon the priority of response of a male-castrate chimpanzee, when paired in the food-getting situation with an unoperated male. They found that, under treatment with male hormone, the dominance of the castrate-male, *i.e.*, his priority of response in the food-getting situation, increased. Treatment with female hormone, on the other hand, produced such a decrease in dominance that the previously dominant castrate became subordinate to his partner. The effect of female hormone on the behavior of this castrate-male thus was seemingly opposite to the effect of female hormone (estrus) on intact females. Several experimentally testable hypotheses, which would explain this differential effect of female sex-hormone level on the socio-sexual behavior of the male-castrate and the intact female, are suggested, and it is indicated that further experimentation is going forward.

In the absence, as yet, of definitive experimental data, the following hypothesis seems to tally best with all observed facts. It posits the existence of two interacting factors—one, hormonal; the other, socio-sexual—operating in the changes of female response-priority with sexual status. The strictly physiological effect of female hormone is to reduce dominance status, as is demonstrated by the experimental findings of Clark and Birch. This action would account for the observed reduction in dominance status, during estrus, of the naturally dominant female in relation to the subordinate male or female partner—a result which has been noted in varied primate types by Zuckerman, Maslow, and Crawford. At the same time, however, a contraposed socio-sexual factor is operative. The perception, by the male, of full female receptivity has, in the chimpanzee at least, a powerful and independent influence on the male's behavior. Relationships in this anthropoid seem to prescribe, as Yerkes has indicated, that the subordinate female, although physiologically less prone to dominance during estrus, shall at such time assume priority of response. Behavior exhibited in this situation indicates that both the male and the female expect such a shift in priority to take place. As already mentioned, male and female patterns of sex response are not sharply demarcated in the primate. Female monkeys and chimpanzees often act like males, especially if they are the dominant animal of a group

or pair. This pervasiveness of the male pattern in female responses would account for observations by Crawford, and by Nowlis, that the naturally dominant female ordinarily grants priority of response to a subordinate female when the latter is in estrus. The dominant animal is, in this situation, responding with the usual socio-sexual custom of the male.

The operation of the two factors outlined would seem to have unquestionable survival value for the primate, in view of Maslow's findings that a dominant monkey, whether male or female, is almost never mounted by a subordinate one, and Yerkes' observations that mating behavior in chimpanzees is mostly unsatisfactory and incomplete where there is a large discrepancy in the dominance status of the consorts. Both the hormonal and the socio-sexual control tend towards producing, in the male and female, a more equal dominance status during the period of mating. If it were not for the former, a naturally dominant female would tend to be sexually inaccessible to a subordinate male. Were it not for the latter, a naturally subordinate female would be so far removed in dominance status from a dominant male that effective mating behavior could hardly occur.

The hypothesis here outlined seems to derive indirect support from an investigation by Nowlis[29] on the relation of degree of hunger to dominance-subordination behavior in the chimpanzee. In this study, it was found that changes in the behavior of the dominant animal have much more effect on the food-taking of the subordinate member of a pair than have great variations in the intensity of drive of the subordinate. On the basis of these findings, it would appear that, in order for strictly hormonal modifications in the female to result in priority of response over a naturally dominant male, it would be necessary that an extremely great increase in dominance accompany an increased level of sex hormone. On the other hand, a much slighter change in the behavior of the dominant male would result in the female's priority of response to the food.

INNATE ORGANIZATION OF MATING PATTERN

Evidence indicates that the mating performance of males of subprimate mammals is innately organized, in reasonably complete form. Beach[2, 3] has shown that rats reared in isolation from weaning until maturity, usually mate as readily and as effectively as an experienced copulator when first placed with a receptive female of the species. As one ascends the evolutionary scale to the infra-human primates,

evidence for the innate organization of the male's mating responses is less conclusive. Factors of learning and social interreactions of varied kinds appear to become increasingly important in determining the mating behavior of monkeys and apes. While there is some indication that the adult male monkey is capable of effective copulation on the occasion of his first contact with a receptive female,[8] this is certainly not true for the chimpanzee. This fact is stressed in the observations of Bingham,[8] and of Yerkes and Elder,[38] on the latter primate type. Although the male may have participated in types of sexual play prior to adolescence, and though the experienced female may attempt to assist him in his efforts, the male is not capable of satisfactory copulatory behavior on his first attempt and requires practice and experience to bring him to a point where he is a skilful copulator, completely acceptable to the female. The female chimpanzee, on the other hand, behaves wholly appropriately and effectively in her first mating experience. It has not been determined if the female chimpanzee raised in isolation would be able to respond adequately to her first sexual contact. It is certain, however, that the process of maturation, plus learning from social contacts and sexual play before puberty, enable the female but not the male chimpanzee to behave appropriately and effectively in the first mating attempt.

BISEXUAL BEHAVIOR

The occurrence of bisexual behavior has been found to be a characteristic of many animal forms. Beach[5,7] has reported many instances of this type of reaction in his series of studies on the sex behavior of rats. Male rats quite often mount other males and execute the entire copulatory pattern. Less frequently, a male rat will assume typical features of the female sexual response when mounted by another male. Female rats, though showing typical sexual receptivity when mounted by males during estrus, frequently respond to receptive females with the execution of the entire masculine copulatory pattern, or of certain elements of it. It is particularly interesting that the most effective stimulus-pattern for calling forth the male copulatory response in female rats is the sexually receptive female: *i.e.*, precisely the same stimulus that is most potent in calling forth this response in the male.

Bisexual behavior is by no means limited to sub-primate mammals. As indicated earlier, such behavior was observed to occur frequently in monkeys. It has also been reported in chimpanzees by Young[39] and by Yerkes.[33] The latter found that associated female chimpanzees

both mounted, and were mounted by, their companions of the same sex. He calls attention to the fact that, when one female mounts another, it is the female in estrus that is mounted. It thus seems that in chimpanzees, as in rats, the receptive female is a potent stimulus for calling forth male copulatory responses in another female.

A study of the sex behavior of infra-human primates and of lower mammals, then, gives indication that the patterns for both masculine and feminine mating-responses are an innate part of the behavior-reaction of both male and female. A thorough understanding of this broad principle, as it operates in animal types so closely related to the human species, should do much to enlighten our socio-cultural attitudes toward so-called "homosexuals" and place this behavior deviation in a proper biological perspective. As I attempted to indicate in an earlier paper,[21] such experimental and observational data should make clear that the either-or type of sexual behavior demanded of man and woman by the mores of Western culture, under threat of severe penalty, is not in line with the trend of sexual adjustment as it has developed throughout biological evolution. In making its present stereotyped restrictions, based upon a dichotomous interpretation of the sex-response, occidental culture overlooks man's biological heritage, his physico-chemical make-up, and his natively endowed response patterns.

In reviewing the experimental studies of sex behavior in infra-human primates, one is struck with the continuity of this behavior with that of lower animal forms. There is nowhere a sharp break from earlier phylogenetic patterns. Though moderately diverse in expression, and though more dependent upon the process of learning, sex remains in primates, as in other animal types, on a manifestly instinctive level. It is uninfluenced by such factors as self-consciousness, or of symbolic processes in the form of mental associations and projections. The marked divergence of the sex behavior of man, when compared with that of other animals, is of particular interest to me because of my association with the phylobiological researches of The Lifwynn Foundation. In man, the biological function of sex has been largely side-tracked by affect-laden images, symbols, and projections. Through an inadvertence in his socio-symbolic evolution, man has in no small measure been betrayed, as it were, by the very symbolic processes that have so greatly enlarged his possibilities of mental adaptation and communication. Because of this *faux pas*, man has now come to mentalize the deepest and most spontaneous needs of his organism. In consequence, sex has, to an immeasurable degree, become lodged in

man's head, and has correspondingly receded from his more basic, instinctive centers of function. The presence of the mother image, of dependence, transference, and sentimentality is due to the intrusion of affecto-symbolic trends that play a major role in the varied dysfunctions in man's socio-sexual behavior.[9,10,11]

The predominant intermixture of partitive, symbolic elements in the sex behavior of the hominid doubtless accounts for the fact that man is the only animal that typically mates without regard to biological fitness for reproduction, and irrespective of the organism's physiological needs. Certainly, the generally accepted conclusion that human sex behavior is uninfluenced by the periodicity of sexual receptiveness and desire would hardly seem warranted, in view of the demonstration of sexual cyclicity in the behavior of a primate so similar to man as the chimpanzee in structure, development, physiological processes, and in many important behavioral aspects. Indeed, on phylogenetic grounds, it seems not unlikely that such a cyclic rhythm would constitute a vital element in healthy human sex behavior, and that it is only because man is dominated by symbolic images and ideas of sex that he has lost touch with the deeper and more basic physiological rhythms which natively control sex reactivity in other animal forms.

LITERATURE CITED

1. **Ball, J., & C. G. Hartman**
1935. Sexual excitability as related to the menstrual cycle in the monkey. Am. J. Obstet. & Gynecol. **29**: 117–119.

2. **Beach, F. A.**
1942. Comparison of copulatory behavior of male rats raised in isolation, cohabitation, and segregation. J. Genet. Psychol. **60**: 121–136.

3. **Beach, F. A.**
1942. Sexual behavior in prepubertal male and female rats treated with gonadal hormones. J. Comp. Psychol. **34**: 285–292.

4. **Beach, F. A.**
1942. Analysis of the stimuli adequate to elicit mating behavior in the sexually inexperienced male rat. J. Comp. Psychol. **33**: 163–207.

5. **Beach, F. A.**
1942. Execution of the complete masculine copulatory pattern by sexually receptive female rats. J. Genet. Psychol. **60**: 137–142.

6. **Beach, F. A.**
1944. Experimental studies on sexual behavior in male mammals. J. Clin. Endocrinol. **4**: 126–134.

7. **Beach, F. A., & Priscilla Rasquin**
1942. Masculine copulatory behavior in intact and castrated female rats. Endocrinol. **31**: 393–409.

8. **Bingham, H. C.**
1928. Sex Development in Apes. Comp. Psychol., Monog. **5**.

9. **Burrow, Trigant**
1917. The genesis and meaning of "homosexuality," and its relation to the problem of introverted mental states. Psychoan. Rev. **4**: 272–284.

10. **Burrow, Trigant**
1918. The origin of the incest-awe. Psychoan. Rev. **5**: 243–254.

11. **Burrow, Trigant**
1937. The Biology of Human Conflict. An Anatomy of Behavior, Individual and Social. 66–103; 347–354. The Macmillan Company. New York.

12. **Carpenter, C. R.**
1934. A Field Study of the Behavior and Social Relations of Howling Monkeys. Comp. Psychol., Monog. **10**.

13. **Carpenter, C. R.**
1941. A Field Study in Siam of the Behavior and Social Relations of the Gibbon (*Hylobates lar*). Comp. Psychol.. Monog. **16**.

14. **Carpenter, C. R.**
1942. Sexual behavior in free ranging rhesus monkeys (*Macaca mulatta*). I. Specimens, procedures, and behavioral characteristics of estrus. J. Comp. Psychol. **33**: 113–142.

15. **Carpenter, C. R.**
1942. Sexual behavior of free ranging rhesus monkeys (*Macaca mulatta*). II. Periodicity of estrus, homosexual, autoerotic, and non-conformist behavior. J. Comp. Psychol. **33**: 143–162.

16. **Clark, George, & H. G. Birch**
1945. Hormonal modifications of social behavior. I. The effect of sex-hormone administration on the social status of a male-castrate chimpanzee. Psychosomat. Med. **7**: 321–329.

17. **Crawford, M. P.**
1940. The relation between social dominance and the menstrual cycle in female chimpanzees. J. Comp. Psychol. **30**: 483–513.

18. **Elder, J. H.**
1936. The time of ovulation in chimpanzees. Yale J. Biol. Med. **10**: 347–364.

19. **Elder, J. H., & R. M. Yerkes**
1936. The sexual cycle in the chimpanzee. Anat. Rec. **67**: 119–143.

20. **Finch, G., R. M. Yerkes, & J. H. Elder**
1937. Bodily electrical potential changes associated with ovulation and early pregnancy in the chimpanzee. Proc. Soc. Exp. Biol. & Med. **37**: 560–563.

21. **Galt, W. E.**
1943. The male-female dichotomy in human behavior. A phylobiological evaluation. Psychiatry **6**: 1–14.

22. **Hamilton, G. V.**
1914. Sexual tendencies of monkeys and baboons. J. Anim. Behav. **4**: 295–318.

23. **Kempf, E. J.**
1917. The social and sexual behavior of infra-human primates with some comparable facts in human behavior. Psychoan. Rev. **4**: 127–154.

24. **Maslow, A. H.**
1936. The role of dominance in the social and sexual behavior of infra-human primates. I. Observations at Vilas Park Zoo. J. Genet. Psychol. **48**: 261–277.

25. **Maslow, A. H.**
1936. The role of dominance in the social and sexual behavior of infra-human primates. III. A theory of sexual behavior. J. Genet. Psychol. **48**: 310–338.

26. **Maslow, A. H.**
1936. The role of dominance in the social and sexual behavior of infra-human primates. IV. The determination of hierarchy in pairs and in a group. J. Genet. Psychol. **49**: 161–198.

27. **Maslow, A. H.**
1942. Self-esteem (dominance-feeling) and sexuality in women. J. Soc. Psychol. **16**: 259–294.

28. **Maslow, A. H., & S. Flanzbaum**
1936. The role of dominance in the social and sexual behavior of infra-human primates. II. The experimental determination of the dominance behavior syndrome. J. Genet. Psychol. **48**: 278–309.

29. **Nowlis, Vincent**
1941. The relation of degree of hunger to competitive interaction in chimpanzee. J. Comp. Psychol. **32**: 91–115.

30. **Nowlis, Vincent**
1942. Sexual status and degree of hunger in chimpanzee competitive interaction. J. Comp. Psychol. **34**: 185–194.

31. **Tinklepaugh, O. L.**
1933. Le comportement sexuel chez les chimpanzés et les singes inférieurs, considéré comme une réaction de substitution consécutive a des troubles émotionnels. J. de Psychol. **30**: 930–954.

32. **Yerkes, R. M.**
1939. Sexual behavior in the chimpanzee. Human Biol. **11**: 78–111.

33. **Yerkes, R. M.**
1939. Social dominance and sexual status in the chimpanzee. Quart. Rev. Biol. **14**: 115–136.

34. **Yerkes, R. M.**
1940. Social behavior of chimpanzees: dominance between mates, in relation to sexual status. J. Comp. Psychol. **30**: 147–186.

35. **Yerkes, R. M.**
1941. Conjugal contrasts among chimpanzees. J. Abnorm. & Soc. Psychol. **36**: 175–199.

36. **Yerkes, R. M., & J. H. Elder**
1936. Oestrus, Receptivity, and Mating in Chimpanzee. Comp. Psychol. Monog. **13**.

37. **Yerkes, R. M., & J. H. Elder**
1936. The sexual and reproductive cycles of chimpanzee. Proc. Nat. Acad. Sci. **22**: 276–283.

38. **Yerkes, R. M., & A. W. Yerkes**
1935. Social behavior in infra-human primates. In: Murchison, C. (Ed.) A Handbook of Social Psychology. 973–1033. Clark University Press. Worcester.

39. **Young, W. C.**
1941. Quart. Rev. Biol. **16**: 135, 311.

40. **Young, W. C., & R. M. Yerkes**
1933. Factors influencing the reproductive cycle in the chimpanzee: the period of adolescent sterility, and related problems. Endocrinol. **33**: 121–154.

41. **Zuckerman, S.**
1930. Proc. Zool. Soc. London 691–754.

42. **Zuckerman, S.**
1932. The Social Life of Monkeys and Apes. Harcourt Brace & Co. New York.

43. **Zuckerman, S.**
1937. Proc. Zool. Soc. London: 107, 315.

44. **Zuckerman, S., & A. S. Parkes**
1932. The menstral cycle of the primates. V. The cycle of the baboon. Proc. Zool. Soc. London: 139–191.

GENERAL DISCUSSION

Dr. Lawrence S. Kubie (*College of Physicians & Surgeons, Columbia University, New York, N. Y.*):

A conference such as this stirs into activity many questions which have been lying latent in my mind. Concerning the patient, careful stalking of an elusive answer to an elusive question, that of the ultimate basis and the ultimate mechanism of sexual need, the work to date has eliminated many candidates for this role, but does not elect any one. By this, I mean that it indicates the existence of a large number of body processes which contribute to sexual activity, but which are not the *sine qua non* of such activity; leaving unanswered the question as to what is the ultimate biochemical basis of sexual need and sexual behavior.

Perhaps I can make this question clearer by indicating, briefly, the state of our knowledge with respect to the other basic instinctual processes, such as respiratory needs, water needs, and nutritional needs. With these, it is clear that, in the so-called resting state of the body, the simultaneous asynchronous processes approximately balance each other. When deprivation occurs, sooner or later this deprivation synchronizes the biochemical processes of the body. This, in turn, creates a generalized tissue need throughout the body, which then must be translated into behavior. In the respiratory sphere, deprivation can last only a few seconds before anoxemia begins to involve all body tissues. Water deprivation takes somewhat longer, and nutritional deprivation still longer, but the basic process is identical in all three.

Is there anything comparable to this in the psychosexual sphere? Are there any biochemical processes which are an essential *sine qua non* for erotic needs? Has deprivation any comparable physiological role in sex? Are local vascular or secretory changes essential, or are all the local changes in the reproductive and genital systems secondary in importance to glandular disturbances (in the pituitary, or elsewhere), or to changes in the central nervous system? Are we in a position, as yet, to answer the positive aspect of these questions?

Dr. Margaret Mead (*American Museum of Natural History, New York, N. Y.*):

It would seem, judging from the material which shows correlations between "intelligence" and efficiency of mating behavior in males, and not in females, that there is a possibility that the wrong units of comparison have been used. If, for mating behavior, we were to substitute the unit, effective reproductive behavior, then we would compare courtship and effective copulation, for males, with courtship, effective conception, gestation, delivery, lactation, and postnatal protective care, for females. We would find, I believe, that the same sorts of relationship existed for females as for males. Choice of the unit called mating behavior is actually based upon an artificial extension, to a two-sex comparison, of the unit which is meaningful for one sex only.

Dr. Bela Mittelmann (*Cornell University Medical College, New York, N. Y.*):

Sexual (genital) behavior can be stimulated, or inhibited, or generally altered, by essentially non-sexual situations and experiences. I would like to cite two observations, one accidental and the other experimental.

I happened to be standing at the elephant's cage in the Central Park Zoo, one day during the War, when the air raid sirens were sounded. The mother elephant and the baby elephant were in the cage. The baby elephant sidled up to the mother. The mother elephant definitely gave the impression, by the posture she assumed, that she was sexually excited. This observation indicates that anxiety-producing stimuli, under certain circumstances, may lead to sexual excitement.

The other observation is of an experimental nature. Gantt induced an experimental neurosis in a male dog, with the conditioning method of exposing him to various stimuli beyond the point of the animal's ability to differentiate. Among other behavioral and psychosomatic symptoms that the dog developed in the experimental situation, there was, also, the phenomenon of almost continuous erection. Further, when the dog was given the adequate sexual object, a receptive bitch, the dog engaged in sexual intercourse, but the duration proved to be shortened. In other words, the dog developed the equivalent of premature ejaculation. In some instances, the duration was one-third of that shown by this dog before the development of the experimental neurosis.

PHYSIOLOGICAL AND PSYCHOLOGICAL FACTORS IN SEX BEHAVIOR

PART II

HUMAN SEX BEHAVIOR

SEX BEHAVIOR IN THE HUMAN ANIMAL*

By Alfred C. Kinsey
Indiana University, Bloomington, Indiana

The data that are presented here, on human sexual behavior, represent a limited portion of material accumulated in an extensive case history study which has been under way at Indiana University in the past eight years. The project is being supported by Indiana University, and by the Medical Divisions of the National Research Council and the Rockefeller Foundation. The study is based upon first-hand interviews with persons of wide social range, of all ages, and of a diversity of educational, occupational, religious, and rural-urban backgrounds. 10,500 histories are now in hand. The material presented in the present publication is particularly concerned with showing the correlation between human patterns of behavior and the basic biological picture which has been developed by the students of sex behavior in other mammals.

Information concerning the ontogenetic development of human sexual behavior has been particularly scanty, because of the difficulty of obtaining information concerning the early sexual activities of children. The data now available have been drawn, partly, from direct interviewing of younger children down to the age of three, where a special play technique has been developed for obtaining evidence, not only of the child's factual knowledge and of his overt experience, but also of his attitudes on sexual matters. Part of the information has been obtained from memories of adults concerning their own pre-adolescent experience, and part has been obtained from observations made by adults on early childhood behavior. It is now clear that sexual attitudes develop at an early period and considerably in advance of the child's acquirement of specific factual knowledge. Attitudes are subtly shaped by the child's younger associates, and by the reactions of its parents and of other adults. At an early age, the child becomes conscious that there are peculiar social values attached to sexual matters, and that certain types of behavior are accepted, while others are condemned. The adult patterns are first laid down in earliest childhood and, by the onset of adolescence, the individual has acquired the pattern which is peculiar to the particular

* The detailed data on which this paper was based will be included in a whole volume on Sexual Outlet in the Human Male, which is now in the course of preparation.

social level in which he lives. Few persons materially modify their patterns during their subsequent lives. The adult thinks that he has consciously chosen his particular course of action, but, in reality, he has merely accepted the judgments of his fellows and a pattern which is long-standing in his community.

Specifically genital play occurs in two-thirds of the pre-adolescent boys, and in about a fifth as many of the pre-adolescent girls. There is more homosexual than heterosexual play. When the play involves contact with an experienced individual, the pre-adolescent may be aroused to the point of an orgasm which is as intense and as complete as anything in the adult. The work with the other mammals has shown that the bases of a complete sexual response are present at birth, and the data on the human are definitely in accord.

In the human, there are many of the same fundamental differences between male and female sexual development and mature behavior that are found in the lower mammals. The responsiveness of the human male develops earlier than that of the female. By fifteen years of age, 92 per cent of the human males have experienced orgasm, but the female population is 29 years of age before a similar percentage has experienced first orgasm. The spread between male and female is much wider in the human than in other mammals. The increased difference in the human is a product of the cultural restraints which delay female sexual development perhaps ten or twelve years longer than is biologically necessary. There is considerable difference in sexual conditionability of human male and female, which is in accord with the work on lower mammals. Males are aroused by a longer list of psychologic stimuli, more readily conditioned by past experience, and more aroused by anticipation of further experience. Visual stimuli are more significant to the average male.

The sexual activity of the male is continuous from the time of the first orgasm, at least if that orgasm occurs after the onset of adolescence. Over 99 per cent of the males have regular, usually weekly to near-daily, sexual activity, from soon after the onset of adolescence until 45 years of age. Female sexual activity is often discontinuous, with gaps of months and years between periods of sexual activity. For a majority of the females, there is no particular disturbance because of lack of sexual outlet during such periods, whereas all males are disturbed when the continuity of their sexual outlet is broken.

Human society has been considerably concerned about the so-called perversions in human sexual behavior, particularly about mouth-genital contacts, homosexual behavior, and such interspecific

activity as the farm boy has with farm animals. The specific record indicates that all of these activities are far more extensive in human behavior than has been realized, occurring in 40 per cent to nearly 75 per cent of large segments of the normal, socially well-adjusted population. No one has realized their wide occurrence, because the social taboos have kept them from public knowledge. The work on other animals now makes it clear that all of these activities are usual and normal parts of the mammalian pattern of behavior. Their extensive persistence in human behavior, in spite of our English-American legal and cultural codes, and abundant evidence that they would occur even more widely if cultural restraint were removed, bear further testimony to their basic mammalian origins. Psychologic studies of these problems have been misdirected. It is not a question of explaining why such things persist in the human animal, but a question of explaining how so many individuals come to avoid and restrain their biologically normal reactions. Further studies of mammalian sexual behavior should be of prime importance in elucidating many other aspects of human behavior.

ABERRANT SEX BEHAVIOR IN HUMANS

By Morris Herman
Bellevue Hospital, New York, N. Y.

According to Havelock Ellis, sexual activities, entirely and by preference outside the range in which procreation is possible, may fairly be considered abnormal, *i.e.*, they are deviations. In order to remain within the normal range, all variations must, at some point, include the procreative end for which sex exists. Of course, one should substitute heterosexual genital contact for the term, procreation, in this definition, since the latter term excludes the normal use of birth control measures. There are serious difficulties in attempting to divide sexual behavior into the dichotomies, normal and abnormal. In reality, there exist kinds of sexual behavior common to the great majority of individuals, and types of sexual behavior which are different in some form or another. The term, aberrant, which literally means straying or wandering away from a straight path, is a good term to apply to these deviations from the average. All too commonly, physicians and others judge these aberrations in sex behavior with inadequate information by their own standard of what is normal or average. It is most important that the knowledge of the average sex conduct of human beings should be widely disseminated, with particular emphasis on the broad variations encountered. Numerous anthropological investigations of the sex life of primitives, and the sex data of contemporary Americans gathered by Kinsey and his group, are exceedingly valuable sources of such information.

It is evident that the relationship of an individual's sex conduct to the mores of the group is a determining factor in the consideration of its aberrant quality. However, many sex acts, such as mouth-genital contacts, etc., are utilized by many as preliminary to heterosexual genital contact. Are these, too, to be considered aberrant? We think not. Aberrant sex behavior can best be defined as sex activity utilized by preference as an end-point in gratification, despite the opportunity and ready availability of heterosexual genital contact.

The term, perversion, as applied to sexual deviation, is mentioned only to be condemned. It conveys a moral judgment which, unfortunately, is already too deeply ingrained in social thought.

Aberrant sexual behavior may be divided into the following types:

(1) Inhibition of sexual impulses; (2) augmentation of sexual impulses; and (3) deflections of sexual impulses.

Inhibition of the sexual impulse is clinically manifested in many ways. An individual may feel no interest of a sexual nature, make no effort to arouse erotic desire within himself, and develop no physical evidence of arousal. Such a condition may result from organic factors, such as the normal process of aging and some endocrine disorders (pituitary underfunction, hypothyroidism, adrenal cortical disease, and sex gland deficiencies). However, many instances occur which are due to psychological factors. A patient has been studied in whom the development of sexual inhibition was related to marked hostility towards a dominant wife. This man was a shy, passive individual. He succeeded, by means of his sexual inhibition, in inflicting punishment upon his wife, who bitterly resented his lack of sexual interest in her. Such total inhibitions may also be a manifestation of a diffuse hysterical process. Frequently, the inhibitions of the sexual impulse are not total but partial, producing such manifestations as impotence or frigidity, premature ejaculation, and vaginismus.

Impotence and premature ejaculation are generally accompanied by great anxiety and distress. In such instances, there is discord between the man's sexual urges and his ability to execute the sexual act, and he usually complains bitterly about his symptoms. The attitude towards frigidity in women is changing. In the last two decades, frigidity was accepted as the normal sex behavior in women. However, the educational aspects of psychoanalysis have broadcast the viewpoint that every healthy woman should attain climax and orgasm. The result is that many women who would have been content with their incomplete sexual performances, if they were living twenty or forty years ago, are not so today. The physician must preserve a scientific attitude, and keep open the possibility that the absence of orgasm may be biological in some women and not necessarily conditioned by emotional factors. Each case must be judged on its merits, and all the biological and psychological factors must be evaluated.

Impotence and frigidity are frequently the result of unconscious conflicts and anxiety. In the analysis of patients presenting such symptoms, one often found the following mechanisms: unconscious fears of punishment; conflicting loves such as parental fixation, latent homosexuality, and narcissism; and conflicting hates or aggression projected onto the sexual partner. A patient has been observed who complained of frigidity, and related the fact that, while having sex

relations with her husband, she feels herself reaching climax, but is frustrated in achieving orgasm by the sudden appearance in fantasy of the face of her brother. Further examination revealed the information that she had had sex relations with her brother, and orgasm when she was nine years of age. This was followed by a profound sense of guilt and fear of punishment. These emotional feelings were instrumental in blocking the occurrence of orgasm in her adult life.

Augmentation of sexual activity is known as satyriasis in the male, and nymphomania in the female. Here again, the problem exists of determining whether an individual is possessed of a greater biological sex drive (a constitutional phenomenon), or whether such heightened sexual activity is manifested as a result of displaced unconscious conflicts and anxiety. There occur individuals who are latent homosexuals and overcompensate for these unconscious feelings by excessive heterosexuality, as if to convince themselves that they are not homosexuals by constant repetition of the sex experience. This group is often referred to as Don Juans or Donna Juannas. In some instances, the psychological meaning of increased sexual activity is a drive for power, in which the individual uses his sexuality to ensnare and conquer. He or she wishes to demonstrate to himself or herself superiority and dominance. Sometimes, the unconscious factors that motivate such behavior are found to be marked aggression and hostility towards the opposite sex.

Recently, there has been described an unusual case of nymphomania, in a woman of forty-five years of age. For several years, she had been experiencing greatly increased sexual appetite, even to the extent of having spontaneous orgasms. The causative factor was eventually determined to be a small tumor in the paracentral lobule of the parietal lobe of the brain, which was producing these sexual symptoms by irritation of the area of the topical projection of the genital structures. The symptoms disappeared completely when the tumor was excised. Another similar case has been observed, in which the excessive erotic symptoms appeared to be the result of stimulation from an epileptic focus in the brain. The sexual feelings in this woman occurred suddenly and intensely, sometimes awaking the patient from sleep. There was no psycho-erotic fantasy preceding the sexual arousal. This woman would either masturbate to relieve her sexual tension, or would engage in sexual intercourse, sometimes requiring the services of ten or fifteen men in one night before her sexual tension was relieved.

The sexual impulse is presumed to be directed towards attaining

heterosexual genital contact. However, it frequently strays and is deflected into other erogenous zones, before achieving genital climax. In some individuals, this *deflection* of sexual activity is the end-point in the process. The psychological sexual process is either abridged or deviated, in such a way that some special part of the process, or some object or action normally on its margin, becomes the chief focus of attention. What is to the normal lover of secondary importance, or even indifferent, thus becomes of primary importance, and may properly be said to be the symbol of the whole sexual process. There are recognized such deflections of sexual conduct as mouth-genital contact, anal intercourse, masturbation, fetishism, sadism and masochism, exhibitionism, voyeurism, and homosexuality. The literature on these topics is enormous, but only a few observations can be made at this time.

Masturbation, as is well known, occurs in a very high percentage of individuals during their early life. This is a normal phase of development. However, in some adults, despite opportunities, masturbation replaces genital contact as a source of gratification, and so becomes aberrant sex behavior. This is found, frequently, in shy, rather introverted characters, commonly referred to as schizoid. It occurs also in neurotic individuals who are unable to achieve genital sexuality. It was so common a manifestation in schizophrenics that the older literature described masturbation as an etiological factor in the production of this psychosis. The masturbation here prepresents a regression to narcissism. The following case illustrates the necessity for evaluating sexual behavior in the setting of an individual's personality: A highly intelligent nineteen-year-old boy, a college graduate, stated that he was nervous and fatigued, and he believed that these symptoms were due to his masturbation. He was aware of the fact that masturbation occurs widely, normally without any ill-effects. But he argued that his body reacted differently from those of other people. This boy was a schizoid character, who expressed his narcissism not only in masturbation symptoms, but also in fantasy formation and projection symptoms, as a result of which he believed that others could recognize his masturbatory activity by changes in the rate of growth of his facial hair.

In a general way, deflected sexuality may be considered infantile sexuality. The cause and nature of this substitution for adult sexuality constitute the main problem. This may be due either to an arrested development, or to a regression. Since the aims of a deflected sexuality are similar to those of infantile sexuality, the possibility for

every human being to develop sexual aberrations, under certain circumstances, is rooted in the fact that he was once a child. In scanning the literature, it is evident that, in different societies and different cultures, these deflected sexual activities have been differently regarded than they are in our present culture. Under varied conditions, such as enforced abstinence in prison, the state of the rejected lover, during alcoholism or drug intoxication, and in the course of psychotic illness, a person may temporarily utilize non-heterosexual genital techniques to obtain sexual gratification. This fluidity of response should make one pause before drawing sharp lines between the normal and the aberrant.

The difficulty of assaying homosexual conduct is a case in point. Freud stated, "I have never yet come through a single psychoanalysis of a man or woman without having to take into account a very considerable current of homosexuality." In practice, we encounter homosexuals who have shown these manifestations since very early in childhood. This gives the impression that the predominant causative factor in these patients is constitutional or biological. These individuals usually are exclusively homosexual. Another group appears to develop following the transition stage of adolescence, in which homosexual conduct occurs, normally, in a high proportion. These individuals generally show, as adults, various admixtures of homo- and heterosexuality. Still another type consists of persons who have been able to repress their strong homosexual tendencies, only to have them produce certain personality traits such as excessive shyness with members of the same sex, Don Juan conduct, or an inability to achieve satisfactory social and sexual adjustment with the opposite sex. The psychiatrist has the opportunity of seeing these mechanisms in operation in the analysis of character disorders, and in the critical observation of the symptoms expressed in the psychoses and neuroses.

The problem of constitutional factors in homosexuality, as shown by physical deviations, has been discussed by Henry and Galbraith, who concluded that homosexual patients show constitutional deviations from the general average. These findings have been contradicted by the work of others. Moreover, one sees individuals with marked secondary sexual characteristics of the opposite sex, and even with varying degrees of pseudo-hermaphroditism, who are without demonstrable homosexual leanings. An interesting case is reported of a person raised as a woman, who had male pseudo-hermaphroditism. Despite the presence of testes and a predominantly male physique, but only a rudimentary vagina, this person had the female psycho-

logical and erotic attitude towards men. It is evident, from such data, that evironmental conditioning is of paramount importance in the choice of sexual object.

When we study social attitudes and the law in relation to sex offenders, we come to a field which is full of challenging problems. Here, law and medical science do not see eye to eye. The law entertains the general attitude that the child is asexual; that any sex behavior on the part of the child is necessarily the fault of someone else; and that the child is always an unwilling victim in any sex episode. This is not true. In some cases, the child is the aggressor. A woman patient related that she began seducing men when she was six years of age. Our society assumes that girls below the age of eighteen are without sufficient will to assent properly to sexual experience. Intercourse with such a female, not one's wife, can result in the arrest of the man for the crime of statutory rape, even though this was not against the desire of the girl. Our society is tremendously concerned with protecting girls from sexual advances by men, but is comparatively unconcerned about protecting boys from sex advances of older women. Many men have been arrested with the charge of indecent exposure of their private parts. Women, however, are almost never arrested on such a charge, even though it is known that such exposure has taken place. At most, they may be charged with disorderly conduct.

Despite the large amount of newspaper publicity generally given to sexual crimes, its incidence in relation to all crime is very low, about 4 per cent of all serious crimes against persons and property. There are about 1,500 arrests, annually, for sex offenses in New York City. More and more, we have to realize that it is not so much the sexual behavior *per se* that is indicative of criminal behavior, but the other aspects of personality reaction. The sexual habits of individuals charged with sex crime differ very little from those of sexual aberrants in the community who have never come in conflict with the law. It is not the sexuality that should be regarded as criminal, but, as in other crimes, the offenses against persons or property.

The problem of the etiology of aberrant sex behavior has been complicated by the efforts of those who attempt to find a single cause for these disorders. As a matter of fact, no simple formula exists. The aberrations of sex conduct result from numerous complex factors, interwoven in such a way that it is often exceedingly difficult to trace them. Cultural-social influences, genetic factors, endocrine disturbances, central nervous system disease, intelligence, constitu-

tional factors, and developmental psychological processes are all important in the production of these deviations. Throughout this dissertation, there have been references to the effect of these factors in the production of sexual changes. The importance of the central nervous system is illustrated in the regressive forms of sexuality found in senility and frontal lobe disease; in erotomania released by irritation of centers by tumor or epilepsy; and in impotence caused by spinal cord disease.

The personality factors in the production of sex aberrations provide a wide field for study and research. In the psychoses, we frequently encounter regressions to infantile sexual traits, either actual or symbolic, such as exhibitionistic masturbation, oral and anal activity, homosexuality, and sadomasochism. In the neuroses, too, there are manifestations of sexual problems in a disguised symbolic fashion. Many sexual aberrations in the adult are the result of conflicts in early life, identical with those of the neurotic. However, they have been elaborated in a different way. In those deviations of personality referred to as psychopathic states, sexual aberration is often present as a result of fixation at some early level of infantile development. Genital sexuality was never achieved.

In conclusion, the psychiatrist brings to the problem of sexual aberrations a technique for the evaluation of personality factors, and a knowledge of emotional disorders from which he can draw comparisons and conclusions. Above all, the psychiatrist, from his clinical experience, is impressed with the variability of human nature. He is aware that, in his dealings with human beings, sexual problems need much more study from all angles. Finally, the psychiatrist must evaluate the theories of sexual conduct and misconduct in the laboratory of actual life.

SEX AND CULTURE*

By Gregory Bateson
John Simon Guggenheim Fellow

It is certainly too early to try to introduce rigor into those anthropological hypotheses which mention sex as a causal factor, or which seek to explain the diversities of sexual behavior by referring to the cultural milieu. Even the word, *sex*, is used by us in a series of different senses, varying from observable and definable copulatory behaviors to a hypothetical drive, or drives, which are believed to influence a very wide and undefinable category of behaviors. It is even doubtful whether we should, at this time, attempt to sharpen our definitions, and not rather wait until some clarity begins to appear, as we amass more data. Definitions and abstractions are, after all, only "right" or "wrong" in so far as they form part of hypotheses which experience can test. Such hypotheses as we have today, relating sex to culture, are still so vague that very much more exploratory work will be needed before the abstractions involved can be sharply defined.

There is, however, a serious drawback to such a *laissez faire* attitude towards theory. It is, unfortunately, easy to construct hypotheses with vague concepts, and such hypotheses are usually impossible to prove or to disprove. The current theories of personality and character formation already contain an excessive number of parentheses (compensation, bisexuality, etc.), any one of which can be invoked to explain why behavior in a given case does not conform to hypothetical expectation. This building-up of parenthetical variables has reached such a point that today it is almost unkind to demand of any theorist. "What conceivable fact could disprove your hypothesis?"

There are, however, two possible approaches which may be of use. not to introduce rigorous hypotheses before the science is ready for them, but rather to suggest the sorts of question which we ought to be asking; and to delimit the orders of hypothesis to which we should look forward.

The first of these approaches will only be mentioned at this stage.

* The original paper as read before the Conference was illustrated by a film of Balinese trance behavior and by a series of Balinese carvings. These materials were used to demonstrate the wide ramifications in human behavior of effects which are in part traceable to 'sex.' The present summary examines more fully than was possible at the Conference the theoretical implications of the sorts of data which were then presented.

It consists in asking metascientific questions about that order of hypothesis which would relate a concept or set of phenomena derived from one scientific field (physiology) to concepts and phenomena derived from another field (cultural anthropology). We are attempting to argue from a narrower sphere of relevance, the individual's internal environment, to a wider sphere which includes almost the whole of human behavior and the external environment. All such transitions from a narrower to a wider sphere of relevance are known to be fraught with difficulty, and we may expect *a priori* that very simple alterations in the narrower sphere will be represented by excessively complex changes in the wider. A small change in atomic structure may denote a total change at the molecular level. Similarly, even so simple a matter as a difference in physical stature might determine very complex differences of culture or society. Physiological sex is known to have causally powerful and complex ramifications within the individual, and, *a priori*, we may expect the social and cultural ramifications of this set of phenomena to be so complex that "sex" will almost cease to be a useful category for the ordering of phenomena at this wider level. Indeed, we know already that those social extensions of "sex" which anthropologists call the "family" and the "kinship system" are crucial to the whole of culture, in the sense that all behavior can be related back to these concepts, just as the same whole can be related back to hunger and the economics of food.[1, 2, 3] This fact, that the effects of any phenomenon within the narrow sphere ramify throughout the *whole* of the wider sphere, indicates that we may not make much headway in attempting to trace the manifold cultural expressions of physiological sex. It is possible, however, that we might make advances by an inverse approach: that, from cultural data, we might be able to derive hypotheses about the narrower physiological sphere. This inverse procedure has an advantage, in that our hypotheses are the more likely to be subject to experimental testing.

The second approach to hypotheses which will relate sex and culture, consists in asking what sorts of data anthropologists do, in fact, collect. This can be followed up with the question: "What

[1] **Malinowski, B.** The Sexual Life of Savages. George Routledge & Sons, Ltd. London. 1929.

[2] **Malinowski, B.** Coral Gardens and Their Magic. 2 vols. George Allen and Unwin, Ltd. London. 1935.

[3] **Richards, Audrey.** Hunger and Work in a Savage Tribe. George Routledge & Sons, Ltd. London. 1932.

types of verifiable hypothesis can be suggested or tested by data of this kind?"

Actually, there appears to be considerable confusion among other scientists, and among anthropologists themselves, about the nature of the data with which the cultural anthropologist works. Therefore, this matter must be made categorically clear. We too often think that the abstractions which we draw are a part of the data from which they are drawn and regard ourselves as studying "culture," or "social organization," or "diffusion," or "religion," or "sex." The creatures which we study are talking mammals and, whether they be natives of New York or of New Guinea, their talk is filled with abstract terms. Thus, we easily fall into the fallacy of assigning a false concreteness to these same abstractions. It is, therefore, salutary, at times, to leave all these abstractions aside for the moment and look at the actual objective data from which all the abstractions are drawn.

There are, I believe, only three types of data in cultural anthropology:

(1) *An identified individual in such-and-such a recorded context said such-and-such, and was heard by the anthropologist.* More than half of all our data take this form, and our main effort in fieldwork goes into the astonishingly difficult task of collecting such items. We do not always succeed, for various reasons. Sometimes, the individual is imperfectly identified. We may have insufficient information about his past experience and position in the kinship system and social organization. Still more often, we may have only an incomplete understanding of the context in which he spoke. But this remains our ideal type of datum.

(2) *An identified individual in such-and-such a recorded context was seen by the anthropologist to do so-and-so.* Here again, the ideal record is not always complete. The identification of the individual and the recording of the context present the same difficulties as in (1), above. In addition, we face very serious technical difficulties when we attempt to record bodily movements. Even with photographic or cinematic techniques, this is almost impossible, and the record, when obtained, can only with very great difficulty be translated into a verbal form for analysis and publication.

(3) *Artifacts (tools, works of art, books, clothes, boats, weapons, etc.), made and/or used by such-and-such individuals in such-and-such contexts.* These are, in general, the easiest data to collect, and the most difficult to interpret.

There are, at present, no other types of objective data in cultural anthropology.*

From inspection of this list of types of data, certain traps in anthropological deduction appear. The most serious of these is baited with the temptation to confuse verbal with behavioral data. Objectively, we may know that an individual said such-and-such about himself, or about some other individual; but we do not know, objectively, whether what he said is true. The objective fact—the only basis upon which we can build—is that he *said* such-and-such. Whether his statement is true or false, must be immaterial to any hypotheses which we may construct, unless (as sometimes, though rarely, happens) we have other objective data bearing upon the truth or falsity of the original statement. The importance of this point can scarcely be overemphasized when we are considering the validity of hypotheses relating to sex—a matter about which human beings are not only reticent and dishonest, but even totally unable to achieve an objective view of their own behavior or that of others.

An example may help to make clear how the anthropologist must proceed in such a case, and how he may construct hypotheses without assuming the objective truth of the verbal datum. Let us suppose that the anthropologist hears and records, *verbatim*, the statement of a man who claims: "I copulate with my wife *n* times every night." This may be an important objective datum, because within it are implicit numerous indications about the psychology of boasting and the psychological role of sex in that man's life. The next question which the anthropologist asks himself will not be: "Is the man's statement objectively true?" He will rather seek for those data which will enable him to place the man's statement in the cultural setting. He will want to know whether boasting and the use of sex activity for the enhancement of self-esteem are culturally acceptable. He will want to know whether such behavior is felt to be aggressive, and against whom the aggression is probably directed, and so on. He

* In defining the contexts of human behavior, many types of non-anthropological information may be necessary. For this purpose, the field anthropologist may have to borrow from almost any of the other sciences. In practice, the "context" of behavior is usually limited to a narration of those factors which the anthropologist deems important and is able to describe. Merely permissive circumstances are usually omitted. For example, meteorological and geophysical circumstances always play a part in permitting an interview to take place, but these factors will usually not be mentioned in the record, unless the anthropologist believes that the shape and content of the conversation was thereby affected. More serious is the common omission of physiological circumstances: the appetitive state of both subject and interviewer, and so on. It is usually impractical to record these in anthropological fieldwork. However, the related sciences of psychoanalysis and experimental psychology are already benefiting from their closer association with the physiological laboratory, and we should certainly look forward to a time when physiologists will collaborate in anthropological fieldwork and amplify the scope of the available data.

will therefore note, first and foremost, who is present on the occasion, and how these other people react to the boasting. He will look to see whether any of those present are markedly superior or inferior in status to the speaker, and whether any women are present; and he may later try to draw comments from the bystanders, after the original speaker has left the group. But in all this, he will not be trying to verify the truth of the original statement, and he will not, in general, care whether that statement is true. At most, he may carry a little suspended query in the back of his mind, a note that this is something about which he does not know, just to remind himself that every hypothesis suggested by the recorded statement must be so constructed that the truth or untruth of the statement will be irrelevant to the hypothesis.

The cultural anthropologist, in fact, is in the peculiar position of studying mammals which talk, and it is necessary to underline this fact to the minds of those who study less articulate and, therefore, less deceptive creatures. The circumstance that our subjects can talk to us, and to each other, is the great advantage which we have over the animal experimenter. However, it is very important not to abuse this advantage. To avoid such errors, stringent precautions must be observed, and these precautions necessarily limit the nature of the hypotheses which we can construct and verify.

Another peculiarity of the data collected by cultural anthropologists is the extreme complexity of each individual datum. The requirement that each datum include full identification of the individual and description of the context, is perhaps never fully met in practice. The fact remains, however, that a very large number of circumstances are always relevant, in the sense that a small change in any one of them might reverse, or drastically change, the form of the behavior which we are recording. There is, therefore, almost no possibility of handling the data statistically. The contexts, the individuals, and the behaviors are too various for their combinations and permutations to be handled in this way. The unit data of which any sample is composed are too heterogeneous to be legitimately thrown together into a statistical hopper. Moreover, the data are not selected at random, but according to circumstances which are forced upon the anthropologist rather than contrived by him. Anthropological informants, of which we do not use very many, are not a random sample of any population. Rather, they are carefully selected and carefully trained individuals, and the characteristics (accuracy, intelligence, articulacy, special interests, special social status, etc.)

which make a man a good informant are not statistically normal in any population. Moreover, the selection is performed by the informants at least as much as by the anthropologist. The man who is in some way deviant, psychologically, sexually, physically, or by social experience, is more likely to want to talk to the anthropologist, and is likely to withstand such interviews with a minimum of boredom. The normal, non-deviant individual rarely, if ever, becomes a regular informant.*

The fact that our data are not suitable for statistical analysis means that they must be handled in other ways. This can be done, just because the unit datum is so complex. It is not necessary to discover hundreds of specimens of *Archaeopteryx lithographica*, in order to satisfy the scientific world that this creature existed and had a number of phylogenetically significant features. The existing samples, consisting of one nearly perfect skeleton, one imperfect skeleton, and one single feather, are more than sufficient, simply because an *Archaeopteryx* skeleton is a complex object. In the same way, the data of the cultural anthropologist, if they are a valid base for theory, are so because they are complex. "This given complex pattern of events occurred"; and this unique occurrence is one of the bricks which must form the material for our theoretical constructions.

This peculiarity of our data, like the unreliability of verbal statements, is a factor which must limit our hypotheses. Neither the single *Archaeopteryx*, nor any number of single specimens of different species, would suffice to demonstrate whether evolution is a continuous or discontinuous process, nor to answer the many sorts of questions which can only be answered by statistical analysis of random or representative samples. Similarly, anthropological data will not suffice to test hypotheses which would require statistical validation, and we must, therefore, avoid hypotheses of this kind. Ideally, we should concentrate upon those hypotheses of which it may be expected that the single exception will disprove the rule.

Within the limitations outlined above, cultural anthropologists have a vast mass of objective data, directly relevant to sex. We cannot report, of course, anything about frequency or characteristics of "normal" sexual intercourse, because such behavior is only accessible to observation under circumstances so exceptional as to preclude use of the term, normal, *e.g.*, copulation on orgiastic occasions, or specially

* There is some variation among cultures in this respect. Among the Iatmul, where verbal articulacy is highly developed, my informants were certainly more culturally normal than among the Balinese, where verbal skill is rare.

staged for the observer and distorted by his presence, or included in traditional dramatic performances.

On the other hand, we have a large mass of information about native notions and psychological attitudes towards sex behavior, and we have information about how these various native attitudes and notions compare, and fit in, with other ideas which the same natives have on other subjects, such as achievement, sadism, humor, prestige, caste, etc. We know a great deal about the stylization of the differences between the sexes, and the roles of the two sexes in daily life and parenthood. We have collected hundreds of fantasies about copulation, love, homosexuality, incest, and so on. Then again, we have a mass of information on the economic aspects of sex behavior: dowries, bride prices, affinal exchanges, prostitution, etc. And masses of gossip about so-and-so's reputed sexual or courtship behavior, data which reflect on the culturally conventional attitudes towards various types of sexual normality and abnormality, and data on the sanctions which the people carry out (or say that they would carry out) in certain types of deviant circumstances.

The problem is to introduce theoretical order into this confused jumble of data, arbitrarily separated from the remainder of our data by their evident relevance to the concept of "sex" which is derived from the physiological sphere of relevance.

Two generalizations can be drawn from all this material: first, that cultures differ markedly among themselves and, second, that a high degree of consistency obtains among the data on any one culture.*

These two generalizations suggest that the regularities which occur within one society† are of a different order from those which occur within an individual organism, and the matter can, perhaps, best be made clear by discussing this difference. If a biologist were allowed to make an exhaustive study of several different tissues derived from a

* Operationally, it is probably necessary to define a "culture" as an aggregate of collected objective data of the kinds mentioned above. This definition may later be amplified, if necessary, by inclusion of some references to the type of order imposed on the data by the scientist. If, however, we limit ourselves to the minimum operational definition, the demarcation between one "culture" and another will have to be defined in terms of causal integration. If we fall into the error of defining this demarcation in terms of homogeneity, saying that culture **A** shall be separated from culture **B** if the data included under **A** differ markedly from the data included under **B**, then we shall have great difficulty in dealing with the differentiation of occupational, age, and sex groupings within the single culture. Still more serious, the generalization in the text above will be a mere endowing of our data with a characteristic of our own operations. If, on the other hand, we delimit our cultures by saying, "Data shall be assigned to a single culture so long as causal interdependence among the data can be recognized," ignoring for the moment the problem of the delimitation of cultures in the time dimension and the related problems of culture contact, we shall, at least, postpone these troubles.

† Operationally, the "society" may be defined as an aggregate of those individuals actually mentioned in the data which constitute a single culture, *plus* those others about whom data can be presumed to belong to the same aggregate.

single species of animal, and then were presented with another tissue differing from all those which he had so far studied, it would be almost impossible for him to determine whether this last specimen was or was not taken from the same animal species. The cultural anthropologist, on the other hand, if presented first with data upon several sorts of individuals in a given society, will probably be able to recognize that the data referring to other sorts of individuals in that same society do, in fact, have that provenience. Moreover, in their attempts to solve this problem, the biologist and the anthropologist will look for clues of quite different sorts. The biologist will look for characteristics of the cells so basic that they persist even through tissue differentiation. For example, he may attempt to count the chromosomes. The anthropologist, on the other hand, will look first for details of language and other very superficial learned characteristics, and if details of this sort are not available, he will look for more basic patterns and regularities which will be diagnostic of the *acquired* character structure of the individuals. It is significant that the biologists talk about "differentiation" of cells and tissues, while the anthropologists talk about the "acculturation" of individuals. The problem stressed by the biologist is: "How do genetically similar cells become different one from another, and how do they maintain these differences in spite of a homogeneous environment internal to the individual organism?" The problem for the cultural anthropologist, on the other hand, is: "How do individuals, who presumably differ among themselves in innate characteristics, become similar and remain sufficiently similar to understand each other, in spite of very evident differences in individual experience?"

In fact, to the cultural anthropologist, man appears not mainly as a physiological mechanism, nor yet as a creature endowed with instinctive urges and innate patterns of response. He appears to us, above all, as a creature which *learns*. The fact of human flexibility under environmental experience determines the major focus of our scientific attention.

Let us now return to the problems of "sex." If learning is the basic concept for the cultural anthropologist, then we can take a first step in relating "sex" to "culture," by examining the relations between sex and learning, bearing in mind that the phenomena connected with learning are to give us the definition of those sorts of regularity and homogeneity which we observe to be characteristic of each single culture. This definition still remains to be drawn.

To the anthropologist, it appears that all human sexual behavior is, in some degree, learned. The human infant apparently de-

velops, at a very early age, a considerable reflex equipment. Its genitalia are erectile in response to various physical stimuli, and this tumescence is early associated with specifically interpersonal stimuli. This reflex equipment is precocious in the sense that the neural connections are present before the infant has the muscular development necessary to put the whole mechanism to work.

In this, there is nothing peculiar about sexuality, and the same sort of precocity is recognizable in other sorts of behavior. The infant also shows what appear to be inherited reflex arcs for walking, swimming, arboreal suspension, and balance in an upright posture. For all of these activities, it appears that neural connections are established before the muscular development is adequate, and the rudimentary responses, which indicate the existence of these neural connections, fade out. A period of "latency" occurs not only in regard to sexual, but also locomotor behavior.*

Now, the crucial question about sexual latency is: "*Is it learned?*" Is the change from genital responsiveness to unresponsiveness to be ascribed to topological changes in the neural network, induced in that network by impulses which pass through it? The obvious alternative to such a hypothesis would be to ascribe the change in responsiveness to changes in the endocrine system and to hope that a fuller knowledge of maturation will enable us to account for these changes in endocrine balance without again being pushed back upon a theory of changes in the neural network induced by experience.†

From what little we know, it appears to me that we must assume that latency is learned, rather than due to a hypothetical endocrine change. The necessary shift in endocrine balance has not been observed,[4] and it appears, rather, that concentrations of androgens and estrogens in children's urine show a progressive rise through childhood to a peak at puberty. Moreover, the hypothesis that latency is learned will have the additional recommendation that it can

* We have, unfortunately, no comparative data about the occurrence of these various types of latency in different cultures, where different methods of handling, carrying, and exercising the baby occur. Even for sexual latency, the data are very poor, and it is perfectly possible that, in some cultures, the potentiality for genital tumescence does not disappear in childhood. It is, however, to be expected that, in all cultures, there is a period during which the child ceases to show specifically sexual desires directed towards adults.

† A third possibility (that all types of latency are due to topological changes in the neural network, but that these changes are a function of maturation and not brought about by the passage of neural impulses) would also be tenable. This hypothesis will, however, differ somewhat from conventional notions about maturation, in that it must account for the breaking or inhibition of previously existing arcs.

[4] **Neustadt, R., & A. Myerson.** Quantitative sex hormone studies in homosexuality, childhood, and various neuropsychiatric disturbances. Am. J. Psychiat. **97**: 542–551. 1940.

be applied not only to sexual but also to locomotor latency, for which an endocrine theory can less readily be imagined.

Granting that learning probably plays a part in causing latency, the next question must be: "What are the stimuli or contexts which determine this learning?" Here we know, from anthropological data, that marked differences exist between cultures. In American and English cultures, we know that, among adults, even the notion of infantile sexuality is strongly resisted; masturbation of the child by parents or nurses is strongly deprecated (and, therefore, probably accompanied by guilt reactions on the part of the adult, when it occurs); and masturbatory behavior on the part of the child is sharply discouraged. Therefore, for these cultures, we must expect that latency will be induced not only by experience of own muscular insufficiency, but also by positive extinction or inhibition of tumescence. In sharp contrast to this, we know of cultures in which masturbation of the child by the parent is common and not deprecated. Even among these cultures, we may expect sharp differences which will be significant for the character formation of the child. The reason given by Italian peasants for this masturbation of the child is "to put it to sleep", and we may presume that, in Italy, the child is given some sort of sexual climax or other satisfactory experience. In Bali, on the other hand, our observations show that the child is not given satisfaction and, instead of going to sleep, becomes more restive. Indeed, from the mother's evident enjoyment both of the baby's responsiveness and of the temper tantrum which often follows, it would appear that the purpose of the masturbation is rather to wake it up. It is easy to see that the Anglo-Saxon and Balinese systems of handling may induce latency, though very different types of latency in the two areas. What sort of latency, if any, is induced by the Italian system, is not so clear.

These contrasts indicate very clearly that the social contexts which accompany the onset of latency are important for sexual learning. They may determine the individual's attitude towards sexual behavior and the part which these behaviors will play in his character. The matter becomes still more complex when we go on to consider the culturally stylized sexual play, masturbation, and courtship behaviors of the latency period, and the rewards and punishments which determine the role of these behaviors in the individual's character. These experiences will label the sexual patterns as safe or dangerous, approved or disapproved, as important sources of self-esteem and prestige, or as important sources of sensual pleasure.

Still later, the onset of puberty can be seen as a further set of learning and character-forming contexts, and the problems are analogous to those which we discussed in connection with latency. Here, the case for ascribing the change in responsiveness to endocrine factors is perhaps a little stronger, but it is possible that the importance of these factors has been exaggerated. There are, in addition, a host of interpersonal and social factors which push the individual into puberty. The sexual initiative of other persons; the value which the individual himself has been trained to place upon sexual adequacy and sexual conformity; his desire to acquire the respect of his fellows of the same sex and of possible sexual partners; all of these, in addition to his endocrines, may push him towards puberty. Further, the psychology of that puberty, when attained, will be determined by the specific qualities of the latency which has been broken down, and by the dynamics of the beginning of puberty. The individual who is pushed into active sexual life, by his physiological needs, at a period when he still feels that social pressures are on the side of latency, will learn something very different from what is learned by the individual who enters upon sexuality in an attempt to conform to social pressures before he is driven to this by physiological need.

The role played by sexual behavior and experience in determining character structure, and the inverse role of experience in determining sexual behavior, could be elaborated almost *ad infinitum*. To relate human sexual behavior to learning is easy, though at every turn we come upon new problems and new hypotheses requiring data for their verification. However, there is still another order of hypothesis which must be considered before we can be said to have related sexual behavior to culture.

It was suggested, above, that the data on a given culture show an internal regularity, or consistency, not simply deducible from the operation which we performed in defining the limits of the single culture in terms of causal interdependence of all events within this margin. We noted, also, that there is a contrast between physiology and cultural anthropology in that, though both sciences deal with spheres of relevance definable in terms of causal interdependence of events within the sphere, the physiologist is preoccupied with differentiation, while the anthropologist is preoccupied with acculturation. Our next step will be to define more sharply these regularities within the single culture, and to relate them to the theory of learning.

What has so far been said about learning could be deduced from any learning theory, such as that of *association*, which will describe

learning as the setting-up of a classification of perceived objects and events, linked to a classification of responses and to a rudimentary linear value system which will discriminate the pleasant from the unpleasant. Such a system will, for example, account for the simple forms of sexual symbolism. Balinese carvings, for instance, illustrate a large number of types of symbolic distortion of the human body. The breasts may be equated with buttocks; the head may be equated with male genitalia; the mouth may be equated with the vulva; and so on.* All of these distortions can be seen as due to simple associational learning.

We find, however, in our cultural data, something more than this. If we take the data from a given culture and sort them by subject matter, putting all the data which refer to sex in one heap, the data referring to initiation in another, the data referring to death in another, and so on, we get a very remarkable result.[5] We find that similar types of order are recognizable in every heap. We find that, whether we are looking at the sex data, or the initiation data, or the death data, the system of classification of perceived objects and events (the *eidos* of the culture) is still the same. Similarly, if we analyze the heaps of data to obtain the system of linked responses and values (the *ethos*) of the culture, we find that the ethos is the same in each heap. Briefly, it is as if the same sort of person had devised the data in all the heaps.

Two obvious hypotheses which might account for this finding can, I believe, be ruled out. We cannot assume that these ethological and eidological regularities are simply due to innate human characteristics, because very different kinds of ethos and eidos have been analyzed out of different cultures. And we cannot say that the ethological and eidological uniformities are due to the uniform working of peculiarities of the mind of the analyst, because in different cultures the same analyst obtains different results.

It would, I believe, be impossible to deduce these results, the uniformity within one culture and the contrast between cultures, from the simple associational learning theory from which we started.†

However, these ethological and eidological uniformities within the single culture, and the corresponding contrasts between cultures, are

* Specimens of these carvings, collected by the writer, were exhibited at the conference. The collection of about 1200 carvings has been deposited at the American Museum of Natural History, New York, N. Y.

† The logical proof of this assertion is, however, beyond my powers, and probably not feasible until the concepts of *eidos* and *ethos* and, indeed, the whole of *gestalt* psychology have been much more critically defined than is possible today.

[5] **Bateson, G.** Naven. Cambridge University Press. 1936. (This book is an experiment in analyzing a New Guinea culture, on these lines.)

precisely what we would expect if, in addition to the processes postulated in simple learning, there is a carry-over from learning in one context which will influence later behavior in quite different contexts. Various theories of this type have been put forward[6, 7, 8] and, in general, the experimental findings indicate that some such postulate may be necessary even at the animal level. At the human level, the carry-over from one context to others can be demonstrated in the phenomena of "transfer" of learning, and especially in the experimental increase in learning proficiency from one context to another of similar formal structure.[9]

Such a postulated carry-over from one context of learning to another will give us a theoretical system which will permit us to speak of changes in *character*, instead of limiting us to the mere addition or subtraction of associational links. We can very easily see how such a theory would give precision to qualities of the order of "optimism," "pessimism," "fatalism," "initiative," "level of aspiration," and the like, and lead us to expect that qualities of this sort, learned by experience in one sort of context, will be carried over into other contexts of very various types. This, I suggest, is the explanation of the ethological and eidological uniformities characteristic of each human culture.

We are driven, I believe, to conclude that what is learned in contexts associated with sex will be carried over into contexts associated with quite different spheres of life—initiation, death, trade, etc.—and that, *vice versa*, what is learned in these other contexts will be carried over into specifically sexual life.

Such a conclusion will reduce the title of the present paper to nonsense, by indicating that *sex* is scarcely a useful concept for the analysis of human cultures; a conclusion which was foreshadowed in our *a priori* metascientific examination of any attempt to relate phenomena in the physiological sphere of relevance to phenomena in the cultural sphere.

Our excursion into theory has not, however, been fruitless, because it has lent anthropological support to a type of hypothesis connected with learning, and this type of hypothesis is such that it can be tested

[6] **Frank, L. K.** The problems of learning. Psych. Rev. **33**: 329–351. 1926.
[7] **Maier, N. R. F.** The behavior mechanism concerned with problem solving. Psych. Rev. **47**: 43–58. 1940.
[8] **Bateson, G.** Comment on **M. Mead's** The Comparative Study of Culture and the Purposive Cultivation of Democratic Values. In: Science, Philosophy, and Religion, 2nd Symposium. New York, N. Y. 1942.
[9] **Hull, C.** Mathematical-Deductive Theory of Rote Learning. Yale University Press. 1940. (This book gives experimental curves for increase in proficiency in rote learning, but does not deduce this increase from a postulate system.)

and made more precise by further anthropological work, and by laboratory experiments. In addition, we have demonstrated that, in the psychological analysis of anthropological data, it is not useful to classify these data according to the sorts of physiological need to which they appear relevant. It is, however, very rewarding to classify these data according to the formal characteristics of the contexts of behavior. It is important to note that the Balinese baby is subjected to the same formal sequence, both when the mother refuses to respond to its temper tantrum and when she cheats it of sexual climax, and that the mother's behavior in both these contexts is an effect of her own past character-formation, as determined by experiences similar to those to which she is now subjecting the child. From such a beginning, we can go on to look at other types of Balinese data, and recognize that the same formal sequence recurs in certain ceremonials in which young men attack a masked figure representing the Witch. They are powerless against her, and fall into a state of disassociation in which they symbolically turn their aggression against themselves, thus achieving an introverted climax.*

From such a systematic analysis of the contexts of learning and the native interpretations of context which are implicit in cultural data, we may hope to build a formal science of culture.

* For photographs of the ceremonial, see **Bateson, G., & M. Mead.** Balinese Character, a Photographic Analysis. Special Pub., No. 3, N. Y. Acad. Sci. 1942.

GENERAL DISCUSSION

Dr. Lawrence S. Kubie (*College of Physicians & Surgeons, Columbia University, New York, N. Y.*):

Even if Dr. Kinsey's report did nothing more than place the facts of infantile sexuality on an unquestionable statistical basis, it would be a major contribution to our understanding of human development and human culture. Psychiatry and psychology will always be in his debt for this. I only wish that some scientific Supreme Court could order certain reluctant psychiatrists to listen to the recital of these facts, every morning, for a month, and perhaps again before they go to bed at night.

In a parenthetical remark of Dr. Kinsey's, there is an implication from which I would dissent. He implied that, when any behavior pattern is widespread among human beings, we need not undertake any "complicated analytical explanations" of it. I am sure that, on further thought, Dr. Kinsey will agree that this is not true. The physiologist does not feel that he does not have to explain the mechanism of the heartbeat merely because everybody's heart beats. Nor does the epidemiologist dismiss the problem of the common cold merely because everybody catches colds. Universality is not synonymous with normality; and our obligation to explain every variation of sexual conduct, whether heterosexual, homosexual, or any other form, is not lessened in any way by the fact that every form of sexual behavior is widespread.

Perhaps, it is worth while to emphasize here the fact that a great deal of the controversy about analysis, and much of the resistance and hostility to it, has arisen because Freud recognized the universality of deviant sexual impulses even in those who give to these deviant impulses no open expression. Certain aspects of analysis stand or fall on the validity of the hypothesis that deviant instinctual tendencies are present, in latent forms, in every human being, derived from the polymorphous infantile sexuality whose existence Freud postulated, and Kinsey has demonstrated, so brilliantly.

Freud related these basic observations to the development of the neurosis, indicating that it was the conflict between latent trends and the forces opposing these trends that splits the personality asunder and leads to psychopathology. It sometimes seems to me that Freud and our analytical colleagues occasionally forget the importance of the conflicts, and speak or write as though it were the latent trend which is itself abnormal and pathogenic. I believe that it would help to bring into harmony the formulations of the biologist and of the psychoanalyst, if we could agree that it is never the deviant drive, as such, which is abnormal, but rather the conflict which arises out of it, and the compulsive and obligatory quality which may attach itself to the drive. An obsessional furor can attach itself to heterosexual activity just as readily as to masturbation or to homosexuality or any deviant forms of sexual conduct. It is the development of this obsessional furor, plus the phobic exclusion of alternative outlets, which is the mark of abnormality, rather than any specific pattern of sexual behavior.

For these reasons, it would add immensely to the value of the observations made by Dr. Kinsey if we could know more about the psychological and psychopathological setting of these various forms of sexual behavior. To this end, it would be essential to gather statistically adequate random samples of each form of sexual behavior, and that these samples be studied intensively by individual psychiatric interviews and by a battery of psychodiagnostic tests.* This, in turn, would lead to much additional information as to the way in which the same individuals handle instinctual processes other than sex, particularly those having to do with food and fluid intake, with excretion, with exercise and sleep. The addition of such information as this would be of great importance for the knowledge of human nature.

* **Harrower, M. R., & R. R. Steiner.** Large-Scale Rorschach Techniques: 419. C. C. Thomas, Springfield, Ill. 1945.
Rapaport, David. Diagnostic Psychological Testing **1:** 574; **2:** 516. Year Book Publishing Co., Chicago. 1945.

Dr. Eilhard von Domarus (*New York, N. Y.*):

In Dr. Kinsey's excellent presentation of man's sexual life, little, if anything, has been said about the influence of different hereditary backgrounds. Assuming that there are, for instance, cases of homosexuality primarily determined by environmental factors, such as of a boy expected to behave like a girl, there appear to be cases where no such environmental factors can be detected and the homosexual conditioning appears primarily, if not exclusively, due to a specific sex behavior determining chromosomal influence. In the latter case, no approval or disapproval of society would be of causative significance. Dr. Kinsey's opinion, with his wide experience in this field, would be extremely valuable.

Dr. Joseph K. Folsom (*Department of Economics, Sociology, and Anthropology, Vassar College, Poughkeepsie, N. Y.*):

Having followed some of the details of Dr. Kinsey's study, I am inclined to agree with him that his results are a fairly accurate picture of the actual behavior of his subjects. Those who have had closer contact with the research, or have themselves been subjected to it, will better understand what he means by the internal consistency of the results and the improbability of any substantial amount of faking.

Furthermore, I think that it is very important to have the kind of data which Dr. Kinsey is providing, even if they are not accompanied by the kind of analysis on the attitude level demanded by Dr. Bateson. We have never yet had, anywhere, such a volume of material on sex behavior. Most of our data have been gathered through questionnaires. This study has the advantage of a well planned, skilful interview technique, and the interviewers have been carefully trained.

Some question might arise as to whether the selection of the subjects has something to do with their personalities and attitudes, even though every socially and externally definable type of person is represented in adequate numbers in the sample. Yet, it must be noted that many of Dr. Kinsey's subsamples included practically the whole population of certain groups to which he spoke. Furthermore, even if it were true that those who volunteered as subjects for such a research, or attended preliminary meetings where the subject was discussed, have personal behavior characteristics different from the average, this would hardly affect the differences which were found between different categories of persons. Moreover, the total is so large that plenty of any given type of person can be found in it, and there is abundant opportunity for checking suspected biases. Many of the subjects were picked up, of course, by casual individual contacts.

From the beginning, I wished that Dr. Kinsey might secure more data of the attitude type, and it appears that his study has already moved considerably in this direction. I hope that, eventually, he may classify and compare his cases, not only by mechanical categories of age, sex, education, etc., but by cultural categories which the anthropologist has found significant. For example, what are the characteristics of Italians, of members of very liberal, religious or other ideological groups, of persons who belong to the upper class according to Warner's community and prestige definitions, not merely according to mechanical definitions of wealth, educational level, etc.?

I am inclined to think that, in the wealth of data that Dr. Kinsey will eventually have, there will be found answers to most of the questions which the anthropologist might ask. I hope that all these questions will be explored, or the data made available to the anthropological, culture-and-personality schools, which indeed possess unique and valuable concepts which should be used in the analysis.

Dr. George R. Bach (*Psychological Clinic, Kent State University, Kent, Ohio*):

I should like to comment on Dr. Kinsey's remarks concerning the methodological difficulties involved in the interviewing of young children on sexual experiences and attitudes. I should think that the use of standardized projective doll play techniques, such as were used in recently published experiments by me, by Dr. Sears and his associates at Iowa University, and by Dr. Conn, should help to overcome this difficulty in future research.

Some of the results obtained with this technique are pertinent to the present discussion. About 5 per cent of experimentally instigated doll play fantasies of normally adjusted children, ages three to eight, have obvious sexual connotations, even when the experimental situation is kept entirely free from suggestions or questions on the part of the psychologist. This is the average figure, the range being from 0 per cent to about 15 per cent. The percentage of sexual fantasies appears to increase with decrease in mental age, when chronological age is held constant. Also of interest to this discussion, is the finding of very definite sex differences in the doll play fantasies of children. Boys show a greater percentage of sexual and other "taboo fantasies" than do girls, whose fantasies, although more numerous, have a greater percentage of stereotypy and propriety. This difference is noticeable at the age of three and seems to be maintained until early adolescence, beyond which we have not as yet used this technique.

I am inclined to explain these very marked sex differences in terms of early cultural sex-typing and principles of social learning: in other words, cultural influences, such as were brought to our attention by Dr. Bateson's paper.

Mr. Amram Scheinfeld (*New York, N. Y.*):

With regard to the reference just made by Dr. Bach to studies of sex differences in projective doll play, it happens that, by coincidence, a copy of one of these studies was sent to me today, and I have it with me. A statement in this paper is of special significance here, because it emphasizes so strikingly the tendency, in recent years, to assume that where we find early sex differences in behavior or performance, they can hardly be otherwise but culturally or environmentally conditioned. I quote now from the study mentioned, by Pintler, Phillips, and Sears, of the Iowa Child Welfare Research Station:*

> "In two recent studies of projective behavior in pre-school children, one . . . on doll play and the other . . . on Rorschach, indications of sex differences have been found. This has two important implications. First, the clinical interpretation of such behavior must be based on intrasex comparisons; the performance must be viewed against the normative background of the child's own sex rather than against that of all children. Second, a new search must be made for the cultural factors that are presumably responsible for sex differentiation. If marked differences are found as early as ages three and four, it suggests that rather powerful influences must be at work from an extraordinarily early age."

Note those last two sentences. Here, quite plainly, is evidence of the bias which has colored so many similar studies, the assumption, *a priori*, that "cultural factors" are "presumably responsible" for the sex differentiation. And yet, at this conference, we have had considerable evidence of biological factors governing sex differences in behavior in lower animals, suggesting the possibility that sex differences in behavior among human beings might also, at least in part, be influenced by biological factors.

Undoubtedly, a great deal of good has come out of the clarification of the important part played by environment and conditioning in the differentiation of behavior in the human sexes. Nevertheless, I cannot help but feel that there is need for a more objective and less prejudiced approach than we have had in the study of the causative factors for these differences, both among children and adults.

Dr. Emily Hartshorne Mudd (*Marriage Council, Philadelphia, Pennsylvania*):

It is well known to all of us that Dr. Kinsey has stated that he does no counseling and takes no responsibility in aiding in problems of sexual or marital adjustment, during his interviews for his fact-finding survey on human sex behavior. On the other hand, those of us who function in the counseling area have met with varied repercussions among individuals who have been interviewed by Dr. Kinsey or his associates. Such individuals are often helped in adjustment, while sometimes they

* **Pintler, Margaret H., Ruth Phillips, & Robert R. Sears.** Sex differences in the projective doll play of pre-school children. J. Psychol. **21:** 73–80. 1946.

are quite upset. Further discussion, with a trained person, of the reactions of such persons, when negative after being interviewed for Dr. Kinsey's study, often helps to alleviate difficulties. It is my feeling, and that of other professional persons with whom I have discussed this matter, that no scientist can completely disclaim responsibility for the results of his work. This point has been amply illustrated in the recent discovery of the atom bomb. It seems to many of us that, during interviews requested by Dr. Kinsey for the collection of his material, other processes, usually of therapeutic value, inevitably take place. Dr. Kinsey might have some suggestions to offer to those of us who are involved in further interpretation of the repercussions of his work.

Dr. Bela Mittelmann (*Cornell University Medical College, New York, N. Y.*):

The statistical and the psychopathological and case history approach complement each other. However, in some ways, they present contradictions and problems in methodology. Dr. Kinsey's approach is, essentially, a statistical one. He surveys the population and finds a series of gradations, and also a universal or almost universal occurrence of phenomena. He considers all of them as normal variations. Dr. Herman presents the point that the phenomena observed in sick patients, particularly if those phenomena appear and disappear with illness, are to be considered pathological. Or, if those phenomena, being present, distress the patient, and disappear as a result of treatment, those phenomena are also to be considered pathological.

Now, the phenomena with which the statistical approach and the psychopathological and case history approach deal, overlap to a considerable extent. I might give two examples. One is partial or complete frigidity in women. The statistical approach finds a very high percentage of it in the general population. It is not an uncommon observation that a woman who comes for treatment, let us say because of anxiety attacks, or because of depressive reactions or obsessional thoughts, and who also is unable to reach orgasm during intercourse, loses her obvious psychopathological symptoms in the course of psychoanalytic treatment and also develops the ability to reach orgasm, or finds it easier to reach one, and thus the frequency of her orgasm increases. The reverse may happen in the case of very frequent and promiscuous relations. In the course of treatment, the level of sexual drive may diminish, and orgasm is reached during each activity, but the frequency of intercourse diminishes. One could make a construction about the first example that the individual is anxiety- and guilt-ridden, with undercurrent hostility, and therefore incapable of adequate sex relations. The second type of individual constantly has to prove his worth and ability of performance, and also utilizes sex relations to allay discomforts from other sources, such as disappointment, hurt self-esteem, and guilt.

There is no doubt that both methods of approach are essentially valid. However, no adequate synthesis of the two types of results has, as yet, been achieved. The statistical approach can be erroneous, because it does not look to the question of constellation and causation. Every individual has had occasional attacks of common cold, and most people have one or two attacks a year. Yet, there can be no question that the common cold is a pathological state. Or, in some geographical regions, the majority of individuals suffer from pellagra, which obviously does not make pellagra a normal phenomenon. The pathological and case history approach may err in that they consider most of the phenomena occurring in the sick individual as pathological. A comparison with broad distribution of phenomena may be lacking. The ultimate synthesis may be achieved by extensive and detailed individual studies, approaching the scale at which Dr. Kinsey is aiming.

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
VOLUME XLVII, ART. 6. PAGES 665–930
MAY 30, 1947

MUSCULAR CONTRACTION*

BY

ALEXANDER SANDOW, R. S. BEAR, ERNST FISCHER, A. S. GILSON, JR., EUGENE GUTH, C. E. HALL, M. A. JAKUS, S. W. KUFFLER, OTTO MEYERHOF, SEVERO OCHOA, ROBERT W. RAMSEY, F. O. SCHMITT, G. M. SCHOEPFLE, SHIH-CHANG SHEN, H. BURR STEINBACH, AND S. M. WALKER

CONTENTS

PAGE

INTRODUCTION TO THE CONFERENCE. BY ALEXANDER SANDOW..... 667

PART ONE: DYNAMICS

DYNAMICS OF SINGLE MUSCLE FIBERS. BY ROBERT W. RAMSEY..... 675

THE TIME COURSE OF TENSION DEVELOPMENT IN THE MUSCLE RESPONSE. BY A. S. GILSON, JR., G. M. SCHOEPFLE, AND S. M. WALKER..... 697

MUSCULAR CONTRACTION AND RUBBERLIKE ELASTICITY. BY EUGENE GUTH.. 715

MEMBRANE CHANGES DURING EXCITATION AND INHIBITION OF THE CONTRACTILE MECHANISM. BY S. W. KUFFLER..... 767

PART TWO: ULTRASTRUCTURE

BIREFRINGENCE AND ULTRASTRUCTURE OF MUSCLE. BY ERNST FISCHER..... 783

ELECTRON MICROSCOPE AND X-RAY DIFFRACTION STUDIES OF MUSCLE STRUCTURE. BY F. O. SCHMITT, R. S. BEAR, C. E. HALL, AND M. A. JAKUS. .. 799

PART THREE: CHEMISTRY

THE MAIN CHEMICAL PHASES OF THE RECOVERY OF MUSCLE. BY OTTO MEYERHOF. 815

CHEMICAL PROCESSES OF OXIDATIVE RECOVERY. BY SEVERO OCHOA..... 835

PART FOUR: MECHANO-CHEMICAL COUPLING

INTRACELLULAR INORGANIC IONS AND MUSCLE ACTION. BY H. BURR STEINBACH 849

THE INTERACTION OF MYOSIN AND ADENOSINETRIPHOSPHATE. BY SHIH-CHANG SHEN..... 875

LATENCY RELAXATION AND A THEORY OF MUSCULAR MECHANO-CHEMICAL COUPLING. BY ALEXANDER SANDOW.. 895

* This series of papers is the result of a Conference on Muscular Contraction held by the Sections of Biology, and Physics and Chemistry, of The New York Academy of Sciences, February 15 and 16, 1946.

COPYRIGHT 1947

BY

THE NEW YORK ACADEMY OF SCIENCES

INTRODUCTION TO THE CONFERENCE

By Alexander Sandow

Washington Square College of Arts and Science, New York University, New York, N. Y.

In welcoming you to this conference, my function is made particularly pleasant by the presence of so many scientific colleagues from abroad. Science, in its essence, is international and, although its local expression may be a product of prevailing social development and need, the exchange of information and the resultant stimulation of new work, which are the life blood of science, transcend national boundaries. With the return to a world of peace, it is heartening to witness here a step toward the restoration to normal vigor of the pulse of international scientific exchange. On behalf of The New York Academy of Sciences, I therefore warmly welcome to this conference our colleagues from foreign lands: from France, England, China, Belgium, and New Zealand, from Germany, Hungary, Latin America, and Spain; and I express the earnest desire that our deliberations will not only further science, but also help in cementing unbreakably our ties of international amity.

The investigation of muscular contraction has, perhaps, no competitor for its preeminent position as the crossroads of the many various active disciplines that comprise biology. Here we find studies of structure: gross, microscopic, and especially today, ultramicroscopic: and of function in its many subdivisions, such as permeability, metabolism in all its chemical and thermodynamic aspects, bioelectricity, secretion—recalling the behavior of the neuromuscular junction, receptor function—as at the muscle spindle, and, of course, above all functions, because most pertinent and fascinating, contractility itself. Moreover, in all these biological pursuits, a most powerful motivating role is played by the methods of the chemist, the physicist, and the mathematician.

There is no doubt that many of these currents of muscle study have richly developed. Yet, generally speaking, there has been among them a provoking lack of integration. These individual lines of research have coursed through their subject, but they have hardly fused into a common trend of progress. Striking limited exceptions to this state of affairs come to mind. We have the notable attempt of the Meyerhof-Hill school to connect the chemical and the heat changes, and that of the Fenn-Hill investigations to relate the mechanical and the heat outputs of muscular activity. But these syntheses, important as they

are, have been concerned essentially with a kind of thermodynamic cost-accounting of contractile activity, and they have contributed only little to the elucidation of the basic molecular mechanism of the muscular response.

Thus, for example, we are in practically complete ignorance of the chain of events that links stimulation, whether by way of the nervous impulse or the direct electric shock, to the contractile machine. Our close friends, the nerve physiologists, in their own conference sponsored by The New York Academy of Sciences,* amply demonstrated the great strides they have made in exposing the detailed processes bound up with the formation and conduction of the nerve action potential. But, as nerve physiologists, their main concern lies in the events along the nerve fiber and the evidently similar changes at the fiber terminations. The muscle physiologist, however, must go much further Not only is he interested in the muscle action potential—with which, incidentally, he has done little—but he would like to know how the action potential excites contraction. His questions concerning this problem are, at present, answered mainly by a tantalizing void.

Another instance of our lack of knowledge is found in the key problem of the mechanical changes themselves, and the word 'changes' is used advisedly, for it is now increasingly clear that we have to account for relaxation of a muscle as well as contraction. We know that the fibrous protein, myosin, or some complex involving myosin, is the contractile material. In the last analysis, this is the fundamental muscle machine that undergoes mechanical change. Obviously, great advances have been made in the clarification of the structure and certain phases of the elastic behavior of this substance. The techniques of protein chemical, and physico-chemical analysis, the methods involving polarized light, x-ray, and electron-beam probes, and the comparisons of myosin with rubber and with biological fibrous materials in general, have all yielded valuable information. Nevertheless, at present, we have only speculation regarding even the composition of the contractile unit, let alone the means by which these units shorten and relax.

Other hiatuses in our knowledge of muscular motion might be mentioned, but it is time we turned to a brighter aspect of our subject. It had long been the dream of muscle physiologists especially to discover the link that connects the mechanical machine of muscle with the chemical cycles that supply the energy for its activity. The search for this link became all the more pressing as many investigations indicated that

* See "The Physico-Chemical Mechanism of Nerve Activity," by David Nachmansohn *et al.* Ann. N. Y. Acad. Sci. 47(4).

myosin was the machine and adenosinetriphosphate (ATP) its immediate source of energy. Now, in the finding of Engelhardt that myosin extracts have adenosinetriphosphatase activity, the basis for the elucidation of that coupling seems to have been determined. Although there are indications that myosin is not itself adenosinetriphosphatase, it is clear that, in the live muscle, the enzyme ATP-ase is exceedingly closely associated with myosin, probably in the form of an adsorption complex. It is not clear what this may mean for the view, originally presented by Engelhardt, that myosin is actively, *i.e.*, enzymatically, involved in the hydrolysis of ATP. But the studies of Engelhardt with myosin threads, of the Needham group with myosin micellae, and of the Szent-Györgyi school with actomyosin complexes leave no doubt that myosin is mechanically affected when it is in the presence of ATP that is undergoing enzymatic decomposition. In this interaction, we are thus dealing with a true mechano-chemical coupling, for the result of the process is not only that adenosinetriphosphate breaks down, but also that myosin is altered mechanically. Thus, this enzymatic activity obviously provides a sound basis for clarification of many problems of muscular activity, certainly those having to do with the direct energization of myosin for its ultimate contractile activity. Dr. Fenn, in his recent review of contractility in Dr. Höber's book, *Physical Chemistry of Cells and Tissues*, having in mind especially the work of Meyer and Astbury and their followers, speaks of the present as the "myosin era." However, in view of the new work flowing from Engelhardt's discovery, I suggest that we are now in the "myosin-adenosinetriphosphatase" era.

Be that as it may, the disclosure of the mechano-chemical coupling of myosin and ATP presents us with a host of new problems. I indicated earlier how disconnected were the several lines of progress in the study of contraction. Now that mechanics and chemistry have come together, they have done so in a head-on collision, and the immediate effect of this is that of any collision—we are presented with a bewildering, but highly suggestive, miscellany that we must assemble and integrate into a harmonious whole.

Undoubtedly, one of the most urgent tasks before us is to determine whether the mechano-chemical coupling of ATP and myosin occurs concurrently with shortening, or during some part of the recovery from shortening, for example during post-contractile relaxation. In the first case, myosin at rest is in a state of low potential energy, and the excitatory process acts to cause it and ATP to interact, so that it is energized for contraction. In the second case, the recovery processes have

energized myosin and, at rest, it is in a state of high potential energy. When excitation occurs, this energy is explosively liberated, and thus contraction occurs. Evidence may be mustered in support of each of these hypotheses, but, at present, a decision in favor of either to the exclusion of the other is impossible.

I wish to discuss, however, several of the more general features of this problem that may be of some value in guiding our thought and research on, at least, the question of the temporal aspects of mechanochemical coupling. It is interesting to note that the general alternative as to whether contraction or relaxation is the active, *i.e.,* energizing, phase of muscular activity, has a long history. Hermann in his *Handbuch der Physiologie,* published in 1879 (Bd. 1, pp. 249–250), lists references as early as 1832 which contain speculations concerned with this point. The modern discussion on this issue begins with the observation of Fenn, in 1924, of the shortening heat—the Fenn effect—which indicates a contraction coupling, and is continued, by means of different approaches, by Ritchie, in 1933, and Kalckar, in 1941, which suggest relaxation coupling.

The terminology introduced by Ritchie requires special discussion. (We need not be concerned with his conclusions, since they were based on the assumption that the initial reaction in the chemical cycle is the breakdown of creatinephosphate, which is not consonant with present evidence.) Ritchie distinguished the two hypotheses presented above by means of the terms, *Chemical Theory,* for energization concurrent with contraction, and *Physical Theory,* for energization during relaxation. It seems to me that these designations, or any variant of them, such as "chemical activation" or "physical activation," should be discarded. Particularly in the molecular realm of interest to the biologist, the distinction between the physical and the chemical is so blurred that it becomes quite meaningless. Above all would this seem to be true in the behavior of enzymes and, as indicated by Bernal, of the contractile material in which the key reactions may be topochemical, that is, "reactions in which one of the reagents is not a free molecule, but a radical bound in a fixed position on a chain molecule." Thus, whenever myosin is energized, the process might still be chemical, in the topochemical sense. It seems, further, that at least now, when we are ignorant of the detailed molecular processes involved in the mechanochemistry of myosin, our terms should emphasize the phenomenological features that are more directly observable. Thus, the question reduces to the following: Is myosin energized for contraction during the engendering of its contractile activity, or during recovery therefrom?

I therefore submit the suggestion that the alternative types of energization of myosin be referred to, respectively, as *activity energization* and *recovery energization;* or, if we wish to emphasize that, in some way, a coupling of myosin and ATP is involved in the energization process, the terms for adoption be *activity coupling* and *recovery coupling.*

Another term that should be examined is *activation.* This term has, of course, a quite precise meaning in the theory of chemical reaction. But it has been used in the literature of muscle physiology to designate what may be described as the mechanism immediately following excitation that sets off the contractile response, as, for example, in reference to Brown's alpha process. It may be well to retain this physiological conception of activation, but then it should be noted that activation of contraction must occur irrespective of whether activity or recovery energization is true. For, if the latter holds in actuality, something must occur to cause the liberation of the energy stored in the myosin. Or, if activity energization is the real process, then there must be some process that permits the previously non-interacting myosin and ATP to interact. In either case, the loosely speaking "tripping-off" action would be designated activation. Thus, activation would refer to some process that temporally intervenes between excitation and contraction, and, generally speaking, it might be distinct from the direct transfer of energy to myosin.

On the other hand, it may be desirable to use the word, *activation,* in the sense of its precise meaning in chemical reaction theory. Then, whenever myosin is raised to a higher level of potential energy, it undergoes activation, and thus we might distinguish between activity and recovery energization by means, respectively, of the terms *activity activation* and *recovery activation.* If this kind of terminology be preferable, then the physiological application of the word, activation, should be dropped, and some other term suggested. It may be noted, however, that, in the case of myosin energized concurrently with the initiation of contraction, the possibility exists that what we referred to before as physiological activation may have its molecular basis in chemical reaction activation.

It is of interest to discuss whether relaxation following shortening is active or passive. The general viewpoint, expressed, for example, in Dr. Fenn's review, is that, if energization of myosin does not occur during the relaxation period, then relaxation is passive. This is not necessarily so. For, even if activity energization occurs, relaxation may still be not passive, in that it depends on some chemical change asso-

ciated with myosin that is relaxing. From what is known of relaxation heat, it would seem likely that this reaction is not exergonic, or only slightly so. However, it may be that some thermoneutral reaction is involved in relaxation, such as the dismutation of adenosinediphosphate to ATP and adenylic acid, or ADP + creatinephosphate → ATP + creatine. Nor must the possibility be ignored that, while energy is needed for relaxation, it need not be stored in the contractile substance for subsequent contraction.

Finally, it should be pointed out that, in the discussion above, an either-or attitude to the question of activity *versus* recovery energization has been adopted. The facts of the matter may not justify this stand. On *a priori* grounds, it seems just as reasonable to presume that another possibility exists, namely, energization during both activity and recovery periods.

I have gone into considerable detail in regard to the general questions of mechano-chemical coupling, since, as will be seen in the ensuing papers, muscle physiology is becoming increasingly concerned with this new and outstanding problem. Furthermore, clarification of terminology in this field is necessary, since that which has sprung up is varied and inconsistent. Certainly, one purpose of a conference such as ours should be the establishment of some common ground concerning at least the basic problems that confront us. Even standardization of terminology will help in achieving a unified viewpoint.

Although mechano-chemical coupling is a recurrent topic in a good many of the papers of this conference, and it is, furthermore, the subject of the entire PART IV, other main fields of muscle study, dynamics (including electrodynamics), ultrastructure, and the chemistry of recovery are discussed in corresponding separate parts. It is to be hoped that our conference will serve not only to advance these various phases of the study of muscle, but that our mutual efforts will also result in some degree of fusion of these rather disparate lines of research into a coordinated attempt to elucidate the fundamental mechanism of contraction. Such a development has been greatly facilitated through the sponsorship of this conference by The New York Academy of Sciences. It is a pleasure to acknowledge our indebtedness to the Academy, and to thank it and its various workers for all that has been done to bring us together to discuss and forward our knowledge of muscular activity. Finally, I wish to extend personal thanks to Doctors Brown, Meyerhof, Ochoa, and Schmitt, whose suggestions and advice were invaluable in the general work of organizing the program of the conference.

MUSCULAR CONTRACTION

PART I

DYNAMICS

DYNAMICS OF SINGLE MUSCLE FIBERS

By Robert W. Ramsey

Medical College of Virginia, Richmond, Virginia

In this paper, it is not proposed to give a résumé of all the diverse phenomena that now have been investigated on single muscle fibers. Instead, some of the simple mechanical properties of uninjured single muscle fibers, in both the resting state and the state of maintained activity (tetanus), will be discussed from the point of view of the influence of the results on current concepts of the muscle machine and its energetic cycle. A more complete and exact discussion was published elsewhere.[1] To a large extent, therefore, discussion of the events between the application of a stimulus and the initiation of contraction will be omitted.

It is generally acknowledged that myosin forms the structural basis of the muscle machine, and that it is closely linked to a chemical reaction that involves a large energy turnover (the breakdown of adenosinetriphosphate).[2] The precise manner of shortening of this protein, whether it is accomplished at the expense of an increase of its entropy or a decrease in its internal energy, and the cycle of the energetic changes, are all matters of dispute for which the mechanics and dynamics observed in single muscle fibers have something to offer.

One simple mechanical fact that necessarily is involved in discussions of latency relaxation, birefringence, the kind of machine (entropy or internal energy), release of resting energy, and so forth, is that of the resting tension exhibited on stretch of a single muscle fiber. Is the equilibrium resting tension due to the muscle substance or to the sarcolemma? Three independent studies germane to this question are those of Sichel,[3] Asmussen,[4] and Ramsey and Street.[5, 6] Buchthal's[7] unsound reasoning on this question will not be discussed. All four investigators would probably agree that, for small stretches, an intact fiber obeys Hooke's Law (as Sichel originally claimed), but that for greater stretches this is no longer true. Sichel maintains that much or most of this resting tension is due to tension exerted by the muscle substance, whereas Ramsey and Street believe they have demonstrated that most of the resting tension is due to the sarcolemma (Asmussen did not measure the tension of the sarcolemma alone). Though these results appear irreconcilable, they may not be, if the experimental conditions of each investigator are considered. Ramsey and Street inves-

tigated intact fibers and then later the sarcolemma of the same fibers, over a considerable range of stretches (rest length to twice rest length). For large stretches, the tensions existing were of the order of a hundred milligrams or more. They were registered by optical magnification of a relatively stiff quartz cantilever. The limiting factor in the accuracy of their measurements was the linear measurement of the magnified excursion of the lever. For a 100 milligram load, the magnified image of this excursion was of the order of 10 centimeters and was read to an estimated half a millimeter. At high loads (100 milligrams or more), the accuracy was at least within 1 per cent, but at low loads of the order of 5 milligrams, the error could be as much as a milligram or a percentage error of 20 per cent. With these limitations, it was found that the resting tension of a single muscle fiber was entirely due to the sarcolemma, since any contribution made by the muscle could not have exceeded a tension of a milligram. Since such a fiber can develop 100–300 milligrams of tension on stimulation, this small possible resting tension of muscle substance was felt to be quite unimportant.

To return to Sichel's experiments. He used pieces of fiber impaled by two glass needles and studied the forces over very short ranges of stretch. If it is assumed that part of his results can be explained by tearing of the sarcolemma by the needles that impaled it, his results can be reconciled to those of Ramsey and Street. Therefore, one would conclude, a very minute tension can be exhibited by muscle substance when it is stretched.

Is this possible very small resting tension sufficient to account for the muscle's thermoelastic behavior, or for the great release of metabolic energy[5, 9] or pH change,[10] on stretch of a resting muscle? I, personally, doubt it, and feel that the phenomena must be related to factors other than that of tension. Since length is the other common factor, it is reasonable to inquire what other changes might be associated with length.

If I may be allowed to indulge in a little pure speculation, certain possibilities may be entertained. Consider the matter of the large increase in resting oxygen consumption observed when a muscle is stretched.[8, 9] There is a considerable body of evidence that suggests that the magnitude of the membrane potential may control the oxygen consumption of a tissue (for discussion of this point, cf. Blair[11]). Moreover, Blair[12] has shown that, from an electrical point of view, the muscle membrane may be considered as a condensor shunted by a resistance, and has suggested that, probably, the product of the capacity times the resistance (cr) remains constant on stretch of the fiber.[13] If

so, the membrane potential will fall until recharged by the increased outward diffusion of potassium which is a consequence of the reduced resistance. In a quick stretch, the fall may be sufficient to account for mechanical stimulation. In a slow stretch, the metabolic systems may be thrown into gear by the falling membrane potential or possibly by the increased outward diffusion of potassium.

What sort of data are available to test this possibility? Hegnauer, Fenn, and Cobb[14] measured the oxygen consumption of muscle in different KCl concentrations, and separately measured the fall in injury potential at the same concentrations. From these data, the resting oxygen consumption can be related to the membrane potential. From Blair's hypothesis, the membrane potential can be calculated for any length of the muscle, assuming that it is not restored. From the data of Meyerhof, Gemmill, and Benetato[5] on the oxygen consumption at

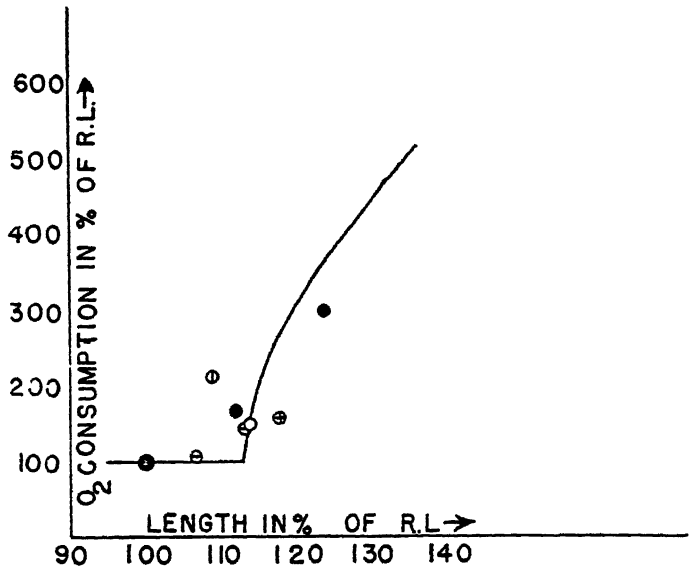


FIGURE 1 Ordinates: oxygen consumption in per cent of that existing at rest length. Abscissa: length of muscle in per cent of rest length. Solid line: a theoretical line calculated from the data of **Hegnauer, Fenn, & Cobb**[14] and the hypothesis of Blair. Circles: three sets of data from **Meyerhof, Gemmill, & Benetato.**[5]

different lengths, the resting oxygen consumption can be related to the length of the muscle. From Blair's hypothesis, and Hegnauer, Fenn, and Cobb's[14] data, let us calculate what the oxygen consumption should be at different lengths of the muscle, and check this theoretical line with Meyerhof's data (see FIGURE 1). When this is done, the agreement is sufficiently close to justify some consideration of the hypothesis.

As another example of more or less pure speculation, one may attempt to rationalize the pH changes undergone on stretching the muscle. According to Dubuisson,[10] stretch of a resting muscle results in an increase in alkalinity.

Suppose that the muscle membrane has a certain number of fixed anionic charges, so that the pH of the surface differs from that of the exterior by a given amount.[15] Moreover, assume that a similar situation exists with respect to the myosin itself. Now if, for any reason such as steric factors or lipoid barriers, these anions could not be discharged, then about them would be a field of cations such as hydrogen ion. In general, the strength of this field would be greatly influenced by the nearness of the anionic charges to one another (though I am not enough of a physical chemist to give formal expression as to how this would vary).

At all events, suppose that a protein fiber, having fixed anionic charges on its surface which could not combine with, but could attract, hydrogen ions within their sphere of action, were suddenly stretched. Under these conditions, an electrode sensitive to hydrogen ions applied to the surface of the fiber and of limited area would indicate an increased alkalinity of the surface, provided that the charges always stayed on the surface (*i.e.*, could not move into the interior of the fiber). Conversely, on shortening, the electrode should indicate a decrease in pH. In principle, this is similar to the assumption Dubuisson[10] made, that the isoelectric point of the myosin is a function of the length. It might be expected that a molecular electrode, such as a redox enzyme attached to the surface of the fiber, would undergo the same pH changes, and thereby influence the rates of oxidative reactions.

If, at its rest-length, it is assumed that the pH of the surface is 0.2 of a pH unit more acid than the bulk phase, then the change in surface pH with change in length of the fiber can be calculated as it is stretched or shortened from the relation:

$$(H^{\tau})_L = (H^+)_{Lo} \frac{\text{Surface area at } Lo}{\text{Surface area at } L}, \tag{1}$$

where $(H^+)_L$ = hydrogen ion concentration of surface at length L,
$(H^+)_{Lo}$ = hydrogen ion concentration of surface at rest-length Lo.

Since the surface area is proportional to the square root of the length of the fiber, EQUATION 1 reduces to:

$$(H^{\tau})_L = \frac{\sqrt{Lo}}{\sqrt{L}} (H^+)_{Lo}. \tag{2}$$

On the stretched side, the pH change calculated from EQUATION 2 can be checked against the data of Dubuisson,[10] and as FIGURE 2 shows, there is reasonable agreement. It should be noted that the highest instantaneous values of the pH change Dubuisson found were used in FIGURE 2. If the means between the instantaneous and equilibrium

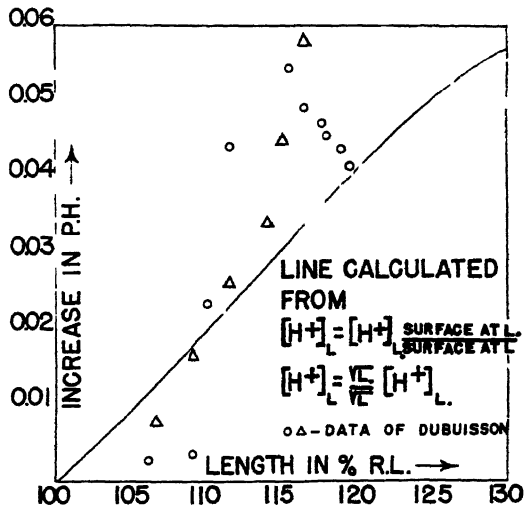


FIGURE 2 Ordinates increase in pH Abscissa length of muscle in per cent of rest length Solid line theoretical calculation from EQUATION 2 of text Triangles and circles two sets of data from **Dubuisson's**[10] Figures 8 and 9

pH changes were used, a somewhat better fit to the theoretical would be obtained. On the shortened side, the pH drops rapidly with shortening and approaches a pH of 6.4 for the greatest possible shortenings of a single muscle fiber, as shown in FIGURE 3.

If this speculation is at all valid, it might provide another possible mechanism whereby a physical change in length of the muscle is enabled to control the rate of release of energetic reactions, since, with respect to oxidative enzymes fixed to the same surface, their potential will be influenced by the pH changes

Another study of the mechanical properties of muscle that has been extended by work on single muscle fibers, has been the establishment of the range and shape of the normal length-tension developed diagram for tetanically stimulated fibers. Ramsey and Street[3] showed that a single muscle fiber developed maximum tension at the rest length (100 per cent), and developed tension reversibly over a range of lengths

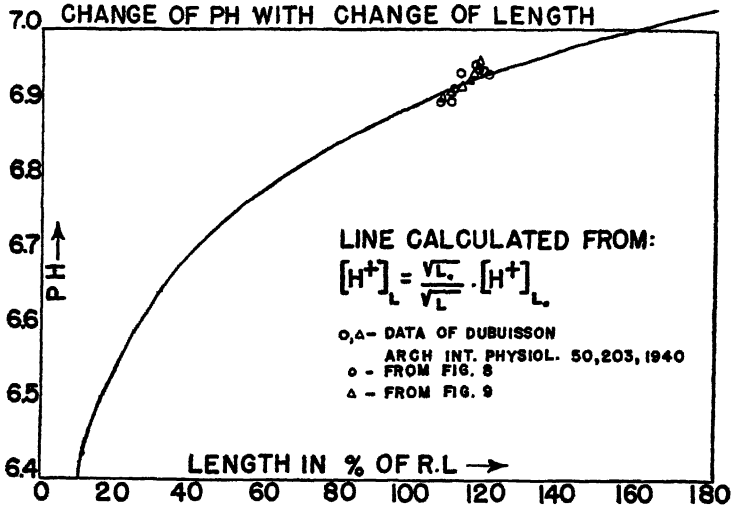


FIGURE 3. Ordinates: pH of surface of muscle. Abscissa: length of muscle in per cent of rest length. Solid line: theoretical line calculated from EQUATION 2. Symbols as in FIGURE 2.

from 67–70 per cent to 200 per cent of rest length.* Further, they found that, if the muscle fiber shortened more than about 30 per cent, there was a striking change in properties of the fiber (the delta state). For convenience, therefore, the normal reversible range will be discussed under the title, "*Normal length-tension diagram,*" and the delta state phenomena under the title, "*The delta state length-tension diagram.*"

NORMAL LENGTH-TENSION DEVELOPED DIAGRAM

Ramsey and Street found that the normal length-tension diagram of single muscle fibers was remarkably constant over a wide variety of conditions. The diagram was the same, irrespective of temperature, size of fiber, amount of previous activity, or season of the year at which the muscles were studied. With this constancy, it is reasonable to try to fit an equation to the data. Using the criteria required to judge the soundness of length-tension experiments, Ramsey[1] concluded that the best existing data for both whole muscle and single muscle fibers gave a length-tension diagram that could be best represented by two intersecting straight lines for the normal reversible range of

* **Buchthal**[17] has recently investigated the length tension diagram of single muscle fibers. In spite of devoting 18 pages to a description of the precision of his apparatus for obtaining absolute measurements, he fails to give anywhere any absolute figures for the tension developed at rest length. From one graph and some of the context of the paper, it would appear that the absolute tensions were so low that he must have been dealing with injured fibers. Buchthal quotes **Asmussen**[16] as having investigated the length tension diagram. If this work was done on single muscle fibers, it would antedate ours. I missed this paper before, but wish to call it to attention, since it was not listed in the bibliography of our earlier paper.

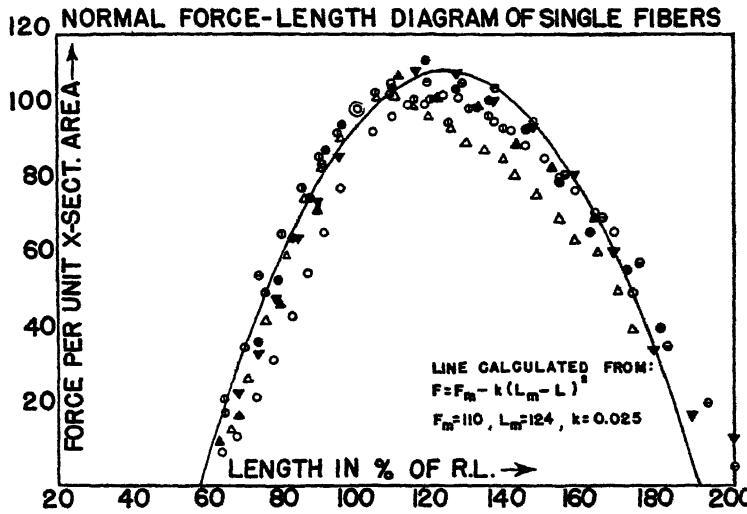


FIGURE 4. Ordinates: force per unit cross-sectional area in arbitrary units. Abscissa: length of muscle in per cent of rest length. Solid line: theoretical force length relation calculated from EQUATION 3. Symbols: five independent sets of data obtained by **Ramsey & Street** on single muscle fibers.

lengths. Unfortunately, few such data exist, and if the attempt is made to contrast the length-tension diagram of skeletal muscle with cardiac muscle, the agreement is only good if the comparison is made with the length-tension diagram of an average experiment. Accordingly, all of the experiments of Ramsey and Street were lumped together, and an average diagram determined. When the tensions are converted to terms of force per unit area, the data of the normal diagram may be expressed in the form of the following equation:

$$Fm - F = k(Lm - L)^2. \tag{3}$$

Here Fm = maximum force per unit area;
Lm = length at which the maximum force is exerted;
F = force per unit area at length L;
k = constant of proportionality.

This equation quite reasonably encompasses the single muscle fiber data, as shown in FIGURE 4, and that of whole muscle as well.

It was of interest to see if cardiac muscle obeyed the same relation. Several sets of suitable data were available in the literature (Doi[17] and Segall and Anrep[18] on the frog heart, and Kozawa's[19] data on the tortoise heart). In all of these experiments, the results were recorded in terms of the isometric pressure developed by the ventricle at various

ventricular fillings. Here, the length of the heart muscle was calculated on the assumption that the contained volume was spherical in shape, while the tension was calculated from the relation between pressure and tension in the walls of a sphere of radius R $\left(T=\frac{PR}{2}\right)$. The rest-length was assumed to be that length that developed maximum tension. With these assumptions, it was found that the heart data fitted the average skeletal muscle diagram extremely well, as shown in FIGURE 5. It also indicated that, normally, the greatest re-

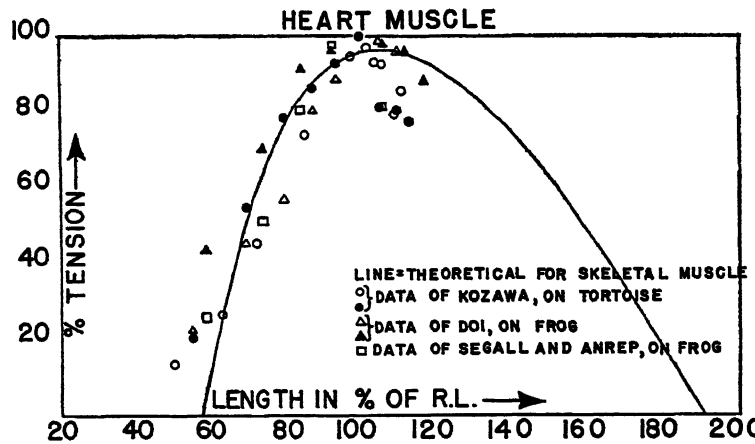


FIGURE 5. Ordinates: tension in per cent of maximum tension at rest length. Solid line: theoretical tension-length relation calculated from the force-length relation to EQUATION 3. Symbols: various heart muscle data.

versible dilation of the heart is reached when the cardiac fibers are at their rest length and exerting maximum tension, for a further increase in length resulted in irreversible loss of tension or return.

Thus, the normal length-tension diagram has great constancy of shape and considerable generality.* Accordingly, any theory of action of the muscle machine must satisfactorily account for this basic mechanical datum. The constancy of the length-tension diagram under all conditions would be expected, if the energetic cycle were such that contraction was accomplished at the expense of energy stored previously in the structure to a saturation point, and that energy from accessory chemical stores restored the system. It is, however, hard to believe that this constancy would be expected if chemical energy had

* **Householder, A. S.,** (Bull. Math. Biophys. 7: 5. 1945) has contributed a theoretical discussion of the length tension diagram.

to be continuously supplied to the structure during contraction. All the complicated kinetics would have to be so ordered that the equilibrium points were the same, irrespective of marked changes of temperature, previous activity, and other varied conditions. It is possible that matters are so ordered, but proponents of the chemical activation theory must give quantitative expression to the length-tension diagram before their views have complete validity.

The normal length-tension diagram, obviously, is of direct concern in any discussion of the molecular basis by which the myosin shortens. In the published literature dealing with molecular structure, a physiologist is hard put to discover, in the diverse proposals advanced by the protein chemists, simple quantitative conclusions that can be tested on living muscle. In this respect, the best of them is that advanced in the numerous papers of Astbury and his collaborators.[20] Astbury suggests that the myosin of muscle at its rest length is in a half-folded (α) form; that, on shortening, this form is converted to a super-contracted form, with a decrease of length from 30 to 40 per cent of the original length; that, on stretch of the muscle, myosin is converted to β myosin, the conversion presumably being complete when the muscle is stretched to twice its rest length. There is little doubt, therefore, that the physiological range of tension development is at least consonant with Astbury's views. Moreover, Astbury implies that there is some peculiarity of structure when isolated myosin has super-contracted some 30 or 40 per cent, since the β x-ray pattern again appears, and this is also the region where single muscle fibers undergo marked changes in properties.

DELTA STATE LENGTH-TENSION DEVELOPED DIAGRAM

The characteristic of the normal fiber that significantly distinguishes it from the delta state fiber is that, upon cessation of the stimulus, the *normal fiber actively relaxes*, whereas the delta state fiber *remains shortened at whatever length it has shortened to.*[5, 1, 21] In other words, the property of *active relaxation is abolished.* It is a very striking change, and occurs wherever the fiber shortens more than about a third of its resting length. The change is permanent and is accompanied by other changes of perhaps equal significance. One of these is that the delta state fiber only develops half or less of the maximum tension that it could before. Another is that it has a new tension developed diagram whose range is now from 20–30 per cent of the original resting length (100 per cent) to 200 per cent. The return

length-tension developed diagram from extended lengths shows enormous hysteresis, as is evident in FIGURE 7. On re-extension, a curve is obtained that approximates the first.

In the delta state, the length-tension diagrams are relatively poorly reproducible on repetition, but invariably more symmetrical when obtained as the fiber is being extended than when it is being shortened If the results for the two best sets of data are utilized, and their tensions at each length converted to forces per unit area, an approximation to the shape of the diagram is obtained with the same formula that holds for the normal, namely:

$$Fm - F = k(Lm - L)^2, \tag{4}$$

where, if both equations have been put in terms of percentage of their own maximal tension, the constant, k, for the delta state, has about two-thirds the value of that for the normal fiber. One such set of data is shown in FIGURE 6, in relation to a line calculated from EQUATION 4.

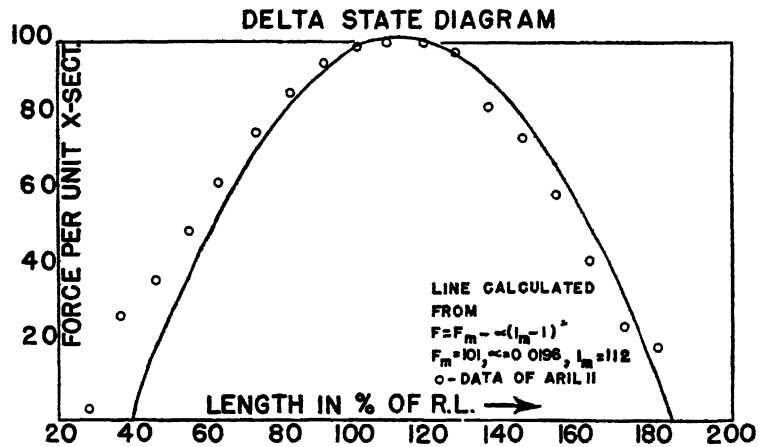


FIGURE 6 Delta state diagram Ordinates force per unit cross-sectional area in arbitrary units Abscissa length in per cent of rest length Solid line calculated from $F = F_m - k(L_m - L)^2$ where $F_m = 101$, $k = 0\ 0196$ $L_m = 112$ Circles single muscle fiber data (April 11).

The delta state has many obvious points of resemblance to contractures and to smooth muscle. In all three, when shortening is extensive, it is very slow. (Even at rest length, the rate of development of tension of the delta state is much slower than in the normal fiber.[5] All are characterized either by failure to relax spontaneously, or else very slow relaxation. Chemical contractures in single muscle fibers

(KCl, caffeine) develop, at their maximum, 25 to 50 per cent of full isometric tetanus tension, just as does the delta state, but, except for acetylcholine, they are not in our experience ever reversible, and the fiber is always permanently damaged. The greatest estimate of the force per unit cross-section developed by smooth muscles is of the order of ¼ that of normal skeletal muscle[22] and thus of an order of magnitude similar to the delta state. Moreover, both the delta state and smooth muscle show the marked hysteresis in their length tension curves when the data are taken by shortening from an extended position, as is well shown in FIGURE 7.

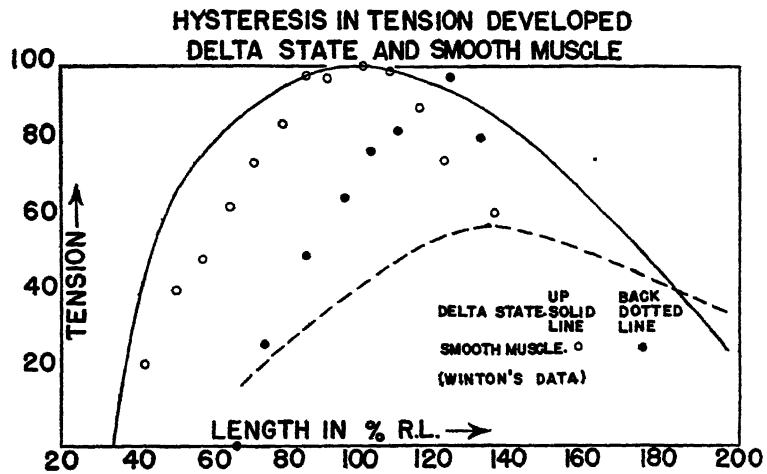


FIGURE 7. Hysteresis in tension developed for a single muscle fiber in the delta state and smooth muscle. Ordinates: tension in per cent of maximum at rest length. Abscissa: length in per cent of rest length. Solid line: tension developed on stimulation by a skeletal muscle fiber in the delta state at various lengths upon extending the fiber from its shortest length. The dotted line represents the tensions developed at various lengths as the fiber shortens from its most extended position. Open and solid circles: similar "up" and "back" data on the retractor penis muscle of the dog from **Winton's**[21] paper.

The hysteresis of the delta state fiber is due to the following phenomena. When an unstimulated extended single muscle fiber is slacked off, it does not immediately take up all of the possible slack, but requires considerable time to reach equilibrium length. Upon stimulation, any of the slack remaining is immediately taken up at the point stimulated and is not distributed over the whole fiber, because the latency between excitation and contraction at a point is much shorter than the conduction time along the fiber. As a result, the cathodal region of the fiber is greatly shortened and the remainder not. In a normal muscle fiber, this shortened region would actively

relax upon cessation of stimulation, but in a delta state fiber it stays shortened. Each time a delta state fiber is slacked off and then stimulated, the shortened region increases in extent. With this increase of the shortened region, there is a correspondingly greater decrease of tension than would have occurred if the fiber were everywhere of uniform dimensions. The result is an enormous amount of hysteresis in the tension developed curves, on return from an extended length. It seems altogether probable that the same type of explanation applies to the hysteresis observed in smooth muscle.

It was, thus, of interest to compare the length-tension diagrams for smooth muscle with that of the delta state. Several sets of data on smooth muscle taken from the literature are compared, the data of Brockelhurst[23] obtained on cat ileum, and those of Winton[24] on the retractor penis of the dog and the guinea-pig uterus. In all these sets of data, the only assumption made was that the length at which the muscle developed maximum tension was equivalent to the rest length.

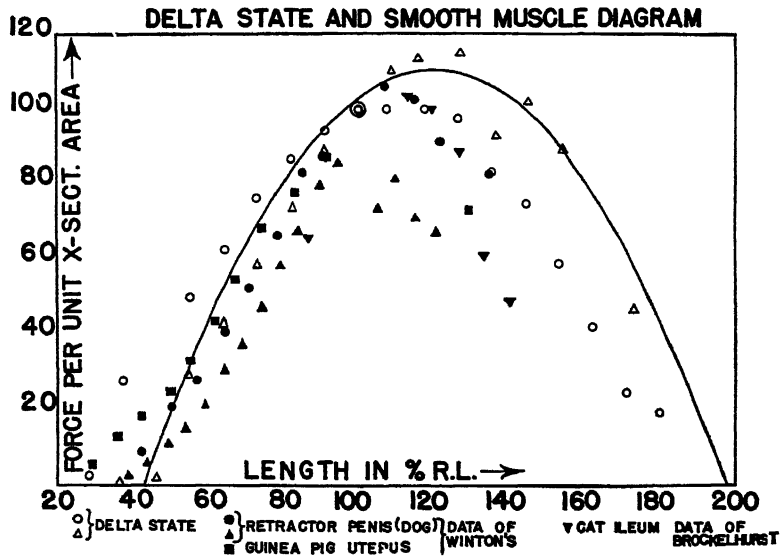


FIGURE 8. Ordinates: force per unit cross-sectional area. Abscissa: length of muscle in per cent of rest length. Symbols. data of single muscle fiber in the delta state, and data obtained on smooth muscles. Solid line. theoretical force-length relation calculated from equation of FIGURE 6.

The above data are plotted in FIGURE 8, and it is clear that the smooth muscle data are as close to the delta state diagrams as they are to each other, or as any two delta state diagrams are to each other. Considering the varied difficulties of the experiments, it is reasonable

to assume that the delta state of skeletal muscle may be closely akin to smooth muscle. If true, it is clear that an understanding of the processes involved in the active relaxation of a normal skeletal muscle and their failure in the delta state will do much to weld the properties of striated and smooth muscle into a graduated continuum.

In this connection, the recent work of Bull[25] is of very great interest. Since this will, no doubt, be discussed elsewhere, it will be sufficient to say that Bull has shown that the length-tension diagram of shortening muscle is very similar to the stress-strain diagram obtained with rubber. He shows that, in the greatly shortened region (delta state), the curve is fitted by the same equation that obtains for non-crystalline rubber, and that the departure from this curve when the normal reversible shortening range of muscle is reached, is similar to the departure of the stress-strain diagram of rubber observed when the rubber begins to crystallize. Obviously, Bull's work opens up a number of interesting possibilities.

TENSION AND RATE OF DEVELOPMENT OF TENSION

The isometric tetanus tension of single muscle fibers varies little with temperature. Dr. Bull[26] recently analyzed some of our unpublished experiments, and he informs me that the tetanus tension increases about 12 per cent for a 10° C. rise in temperature. According to Dr. Bull, this is consistent with the idea that the muscle is an entropy machine. It would be of interest to hear from the physical chemists if it is necessarily inconsistent with an internal energy machine.

Dr. Alan Young and I, independently, made numerous attempts to derive some system of sensible kinetics that would account for the rate of tension development and relaxation in twitches and tetanus records obtained from single muscle fibers. As far as I am concerned, I was completely unsuccessful. Either twitch or tetanus records could be readily fitted by a variety of equations, but attaching any sensible meaning to the constants was another matter.

The difficulty with the twitch analysis is that, in skeletal muscle, the latency between excitation at a point on the fiber and contraction of the same region is much shorter than conduction of the impulse down the fiber. The result is that the region first stimulated contracts and stretches, markedly, the part still unexcited. Later, when the impulse reaches this region, it starts to shorten and stretch the previously shortened region. While this behavior is best seen in

single muscle fibers,[1] it was observed much earlier in whole muscle by Fischer.[27] Even when the fiber is stimulated in several places simultaneously, the effect is quite prominent, and a little reflection will show that the type of equations necessary to deal with such behavior falls into the class known as integral equations (which, incidentally, are quite beyond the borders of my mathematical competence.). Recently, Drs. Gilson, Walker, and Schoepfle[28] have approached the kinetics of muscle, starting with the twitch as the basic pattern. Unless the above phenomenon is taken into account, their analysis is not likely to be as productive as it might otherwise be.

The second great difficulty is the enormous change in the kinetics of both twitch and tetanus with change of length of the muscle. This may not be so apparent in studies of whole muscle, where the range of lengths that can be studied is greatly restricted by the connective tissue, but is very obvious in single muscle fibers, where this restriction is greatly mitigated.

Since I was unsuccessful in fitting the data by any system of fundamental kinetics, I endeavored to work out the simplest empirical equations that would satisfactorily fit the rate of development of tension at all lengths of the muscle. The result was as shown previously,[1] that the rate of development of tension at any length from rest length (Lo) to twice rest length was described by the following equation:

$$T = T_L[1 - Ae^{-\alpha t} - (1 - A)e^{-\beta t}], \tag{5}$$

where T_L = maximum tension at length L,

$$A = \frac{2Lo - L}{Lo}, \quad 1 - A = \frac{L - Lo}{Lo},$$

α and β = constants.

At rest length (Lo), $A = 1$ and $1 - A = 0$. Therefore, the equation becomes:

$$T = T_{Lo}(1 - e^{-\alpha t}). \tag{6}$$

When the muscle is allowed to shorten and then develop tension at the shortened length, the rate of development of tension is described by:

$$T = T_L(1 - e^{-\alpha t}), \tag{7}$$

where α continuously decreases with shortening, starting from the value at Lo and diminishing to the value β at $\frac{2Lo}{3}$. With greater shortenings (below $\frac{2Lo}{3}$), α remains approximately constant and equal to β.

These relations, of course, immediately suggested correlation with Astbury's hypothesis. In the detailed analysis of the assumptions necessary for such a correlation, assuming the validity of Astbury's ideas, it was shown[1] that the principal ones were: that one must assume that the strength of the individual α myosin links differed from one another by a small amount (perhaps because of differing side chain linkages); that in a polar environment, the weaker myosin links would unfold to the β state before the whole had to be stressed to the average critical level; that at twice rest length, and possibly two-thirds rest length, all of the myosin was in β form. With these assumptions, a stretched muscle would have some of its myosin in α form and some in β form, and the proportions of each molecular species would be specified by the length.

In the equations, therefore, if the constant α represents a velocity constant for α myosin, and the constant β that for β myosin, then the coefficients A and $(1 - A)$ would specify the proportions of α and β myosin, respectively, existing at the extended length. Numerically, these coefficients are what would be expected, if Astbury's hypothesis and the above assumptions are correct.

As noted above, the rate of development of tension at shortened lengths between $\frac{2Lo}{3}$ and Lo can be adequately described by a single exponential term whose constant decreases in value with shortening. One suspects that, if the data were better, it would be found that, in this region, the rate of development of tension was given by the same equation as that for the extended lengths, but that the relation of the coefficients to length were different. When the fiber is shortened below $\frac{2Lo}{3}$, the data are poor, but approximately the single exponential term has a constant that is equal to β.

Summarizing, then, the myosin of muscle might be visualized in the following terms: At rest length, it exists in a ½-folded or α state. Upon extension, the α myosin links are extended, one by one, to the β form, each having twice the length and half the cross-sectional area of an α link, so that, at twice rest length, all of the myosin is in β form. At intermediate lengths, the proportions of α and β are specified by the length. A similar situation may be occurring in shortening, resulting in the conversion of the α myosin again to β form, so that, at $\frac{2Lo}{3}$, there are now 3 β links per unit cross-section, instead of the one β link at 200 per cent extension.

Such a model would very accurately account for the physiological data obtained on single muscle fibers. It would account exactly for the range of the reversible length-tension diagram; it would provide a reason for the striking change in properties of a muscle when it shortens more than a third of its length; the shape of the length-tension diagram would be explained if both an α and β myosin link could develop tension, but the β link could only develop about a tenth that of the α; and, finally, it would completely account for the empirical kinetics of the rate of development of tension *at all lengths of the muscle.* In general, this evidence constitutes rather strong physiological support for Astbury's ideas.

The rate of development of tension, preceding and following prolonged activity, is of some interest in relation to the two hypotheses of the energetic cycle. In whole muscle, the rate of development of tension immediately following previous prolonged activity is always slower than it was before activity,[29] though recently this has been questioned.[28] On the activity energization hypothesis, this is to be expected. But, in single muscle fibers, *this slowing never occurs;* the rate of development of tension immediately after prolonged activity is always equal to or faster than what it was before activity; while the only result of the previous activity is greatly to prolong relaxation. The results on single fibers, therefore, are much more in keeping with the recovery energization theory.

In general, fatigue of the contractile system is never observed in single muscle fibers. Whether the fiber was stimulated directly or indirectly through its nerve, the maximum tetanus tension was maintained until excitation failed and the fiber abruptly failed to respond or only responded to every second, third, fourth, or other stimulus.[6, 30] Asmussen[31] found the same phenomenon with indirect stimulation. All of the indirect evidence obtained on whole muscle is consistent with the view that the decrease of tension in fatigue is merely due to the fact that fewer muscle fibers are contracting, and the direct evidence from single muscle fibers confirms that view. This is another bit of evidence which is not easy to reconcile with the activity energization theory, but would be expected on the recovery theory.

THE ENERGETIC CYCLE

In the establishment of an energetic cycle, it is a necessary condition for validity to show that the proposed cycle agrees. in respect both to order of appearance and magnitude of energy changes, with those

exchanges actually observed. In a system of multiple possibilities, however, establishment of such agreement cannot be accepted as proof of validity, but can only establish the claim that the hypothesis can be entertained. In general, the only occasion that will lead to proof of validity in such a system is that which arises when failure to close the cycle at a particular point leads to predictable physical consequences which can be shown do occur.

There can be little question that the most powerful experimental support for the activity energization theory derives from the phenomenon now known as the Fenn effect.[12, 32] Deferring consideration of this phenomenon for a moment, it is clear that Hill's success in relating the mechanical events of muscle contractions to the liberation of energy, both in respect to time and magnitude, has gone far towards establishing the adequacy of this particular cycle. This formulation, however, has two principal defects. One is that no limit is set to the extent to which a muscle can shorten. In so far as the formulation is given, a muscle could always shorten to the same extent, irrespective of load, and this is completely incompatible with the experimental facts, since a definite length-tension diagram exists. The second defect is that, in order to have agreement between the energy exchanges and mechanical events, Hill had to assume that, at any particular load, the velocity of shortening remained constant, irrespective of the extent of shortening. Now it is an experimental fact that the velocity of shortening is not constant, irrespective of the extent of shortening, but diminishes as shortening proceeds. This is clearly evident in Fenn and Marsh's[34] experiments. It is also shown well in the records of free shortening of single muscle fibers.[6] Hill's assumption, therefore, is only valid provided the shortening is restricted and the loads are not too heavy.

To return to the Fenn effect, if the recovery exchanges are ignored for the moment, the extra energy liberated in shortening is equal, or very closely equal, to the work done (Fenn,[32] Hill,[33] Brown[35]). When the muscle exerts tension equal to the load and then begins to shorten, it is the structural basis of the muscle that is bearing the load and doing the shortening, and not the accessory chemical reactions. It is hard to avoid the conclusion, therefore, that it is the particular configuration of the structure that is governing the energy release and that this energy release is occurring in point-to-point fashion, as shortening proceeds, since the extra energy is so perfectly proportioned to the work done. It is difficult for me, therefore, to visualize a

mechanism whereby the energy released is antecedent to the structural change, as it requires to be on the activity activation theory.

When one tries to reconcile the recovery activation theory with the Fenn effect, however, immediate apparent difficulties arise, but these difficulties are almost wholly due to the fact that one now tends to visualize the activated muscle as a stretched spring having properties similar to those in the world of mechanics. The point of importance is that the energy dissipated when a stretched spring reaches its equilibrium point, is independent of the path by which the equilibrium was approached, and *that, at equilibrium, the state of the molecules composing the spring is everywhere identical.* The conviction of some that the Fenn effect can only be explained on the basis of the activity energization theory is due, I believe, to the fact that they are making the implicit assumption that, if a muscle is exerting the same tension at the same length, all of its molecules must necessarily be in the same energetic state: *that is, the muscle is still being visualized as a simple spring, but one whose momentary potential energy is governed by the rate at which energy can be poured into it from the chemical stores.*

But, if a muscle can exert the same tension with two totally different configurations of its molecules, having, therefore, liberated two totally different amounts of energy, then the simple spring idea cannot be entertained. With its abandonment goes the necessity for using the activity activation theory to explain the Fenn effect, *since the only fundamental phenomenon,* the recovery activation theory, requires to explain that the Fenn effect is some mechanism whereby the total energy liberated in shortening is dependent upon the manner in which the shortening takes place.

As shown previously,[1] several simple experiments on single muscle fibers show that this requirement is met. For completeness, they will be reiterated here. If a single muscle fiber is just stretched between supports to its rest length, and is stimulated tetanically, simultaneously at several points along its length, it will develop maximum isometric tension, and the fiber will everywhere present a uniform appearance. Now move the supports nearer to each other by an amount about equal to 15 per cent of the rest length. On this occasion, arrange matters so that the fiber is only stimulated at one end. The moment stimulation begins, the end first stimulated shortens enormously and takes up all or most of the slack before the farther end begins contracting, so that about a third of the fiber is enormously shortened and bulged out, while the remainder has about normal rest length diameter. Under

these circumstances, the tension registered in the *first second or so of the tetanus* may be almost exactly equal to that registered when the fiber was held at rest length. (As the tetanus is continued, the unequal shortening is equalized along the length, and the tension falls to the value expected for a shortening of 15 per cent.) Thus, for at least a second, there can be no question but that the muscle machine exerted the same tension at two totally different configurations of its molecules and with the liberation of two totally different amounts of energy. By suitably arranging the position and number of electrodes, it is possible to have the same fiber shorten to the same extent and exert the same tension, but visibly have two totally different configurations: that is, shortening can be even, or a part of the fiber can shorten very greatly, leaving the remainder in either a stretched or rest length position. Finally, if a fiber held at rest length is stimulated at one end, the same sort of process happens, but it is very much reduced in extent. In other words, whenever a fiber must first develop tension before overall shortening takes place, the effect of the load is to equalize the pattern of shortening, irrespective of where the fiber is stimulated. *The load, therefore, can govern the pattern of shortening or configuration of the molecules and, thus, the energy released on a purely recovery energization basis.*

Since, on a recovery energization basis, the area of the length-tension diagram specifies the total energy available, it necessarily must be directly related to the rate of the release of energy. In order to relate the kinetics to the length-tension diagram, I assumed[1] that the velocity of shortening at any load was proportional to the distance to which the muscle still could shorten, an assumption reasonably justified by the experimental facts. If this is granted, I do not believe that the heat exchanges offer any impediment to acceptance of the recovery activation theory.[1]

Obviously, the above is not a model of the muscle machine, which must be left largely in the hands of the physical chemists. It does demonstrate, however, that the physiological facts can be interpreted on the basis of a recovery energization model, just as well as or better than on an activity energization system.

LITERATURE CITED

1. **Ramsey, R. W.**
 1944. Muscle: Physics. In: Medical Physics, ed. Glasser. Year Book Publishers. Chicago.
2. **Engelhardt, W. A., & M N. Ljubimowa**
 1939. Nature **144**: 668.
3. **Sichel, F. J.**
 1935. J. Cell. & Comp. Physiol. **5**: 21.
4. **Asmussen, E.**
 1939. Quoted by **F. Buchthal & J. Lindhard.** Det. Danske Videnskab. Selskb. Biol. Meddel. **14**: 68–70.
5. **Ramsey, R. W., & S. F. Street**
 1940. J. Cell. & Comp. Physiol. **15**: 11–34.
6. **Ramsey, R. W., & S. F. Street**
 1941. Biol. Symp. **3**: 9–34.
7. **Buchthal, F.**
 1942. Det. Danske Videnskab. Selskb. Biol. Meddel. **17**: 1–138.
8. **Meyerhof, O., C. L. Gemmill, & G. Benetato**
 1932. Biochem. Z. **258**: 371.
9. **Feng, T. P.**
 1932. J. Physiol. **74**: 455.
10. **Dubuisson, M.**
 1940. Arch. Internat. Physiol. **50**: 203.
11. **Blair, H. A.**
 1938. J. Neurophysiol. **1**: 127–143.
12. **Blair, H. A.**
 1941. Biol. Symp. **3**: 51.
13. **Blair, H. A.**
 Personal communication.
14. **Hegnauer, A. H., W. O. Fenn, & D. M. Cobb**
 1934. J. Cell. & Comp. Physiol. **4**: 505.
15. **Davson, H., & J. F. Danielli**
 1943. The Permeability of Natural Membranes. Macmillan. New York.
16. **Asmussen, E.**
 1936. Skand. Arch. Physiol. **74**: 129.
17. **Doi, Y.**
 1920–1921. J. Physiol. **54**: 218.
18. **Segall, H. N., & G. V. Anrep**
 1926. Heart **13**: 61.
19. **Kozawa, S.**
 1915. J. Physiol. **49**: 233.
20. **Astbury, W. T., & S. Dickinson**
 1940. Proc. Roy. Soc. London B **129**: 307–332.
21. **Fenn, W. O.**
 1945. Chapter in: Physical Chemistry of Cells and Tissues, ed. Hoeber. Blakiston. Philadelphia.
22. **Fischer, E.**
 1944. Physiol. Rev. **24**: 467.
23. **Brockelhurst, R. J.**
 1926. J. Physiol. **61**: 275.
24. **Winton, F. R.**
 1926. J. Physiol. **61**: 368.

25. Bull, H. B.
1945. J. Am. Chem. Soc **67:** 2047.
26. Bull, H. B.
Personal communication.
27. Fischer, E.
1926. Pfluger's Arch. **213:** 352.
28. Gilson, A. S., Jr., S. M. Walker, & G. M. Schoepfle
1944. J. Cell. & Comp. Physiol. **24:** 185.
29. Hill, A. V.
1931. Proc. Roy. Soc. London B **109:** 267.
30. Ramsey, R. W., & S. F. Street
1942. Fed. Proc. **1:** 70.
31. Asmussen, E.
1934. Skand. Arch. Physiol. **70:** 233–273.
32. Fenn, W. O.
1923. J. Physiol. **58:** 175–203.
33. Hill, A. V.
1938. Proc. Roy. Soc. London B **126:** 136–195.
34. Fenn, W. O., & B. S. March
1935. J. Physiol. **85:** 277.
35. Brown, D. E. S.
1941. Biol. Symp. **3:** 161.

THE TIME COURSE OF TENSION DEVELOPMENT IN THE MUSCLE RESPONSE

By A. S. Gilson, Jr, G. M. Schoepfle, and S. M. Walker

Department of Physiology, Washington University School of Medicine, St. Louis, Missouri

The time consumed by the single response of a muscle varies greatly for different muscles (FIGURE 1). The sartorius muscle of the frog, a striated muscle, requires a few hundredths of a second (ca. 0.04 sec.) to attain maximum twitch tension, and a few hundredths more to return to resting level. The retractor penis of the turtle, also a striated muscle, attains peak tension in a few tenths of a second (0.4 sec.). Smooth muscle from the intestine or stomach of the turtle requires some 30 seconds to reach peak tension Certain differences in the forms of the records obtained from these various muscles will be discussed later in this paper, but it is apparent that, if the time scales be adjusted, there is striking similarity between the response curves of different muscles.

In an attempt to achieve a more precise description of the twitch response to a single stimulation, records of isometric responses of the frog sartorius muscle were studied (Gilson, Walker, and Schoepfle, 1944). The records were obtained by means of an optical myograph and a magnification of movement up to 400x, so that a twitch record showing 4 cm. excursion might involve an overall shortening at the attachment of the muscle to the level of only 0.1 mm. Selected records of the simplest form obtainable showed consistent agreement with curves derived from mathematical treatment, using what appeared to be reasonable assumptions concerning the processes involved in the muscle response.

For purposes of mathematical treatment, the muscle has been considered to be an elastic body having mass and viscosity, attached to the spring lever which possesses its own elastic characteristics. For the stretched resting muscle, certain balanced forces exist. When the stimulated muscle responds and exerts further tension against the lever by shortening, however slightly, it does so by development of

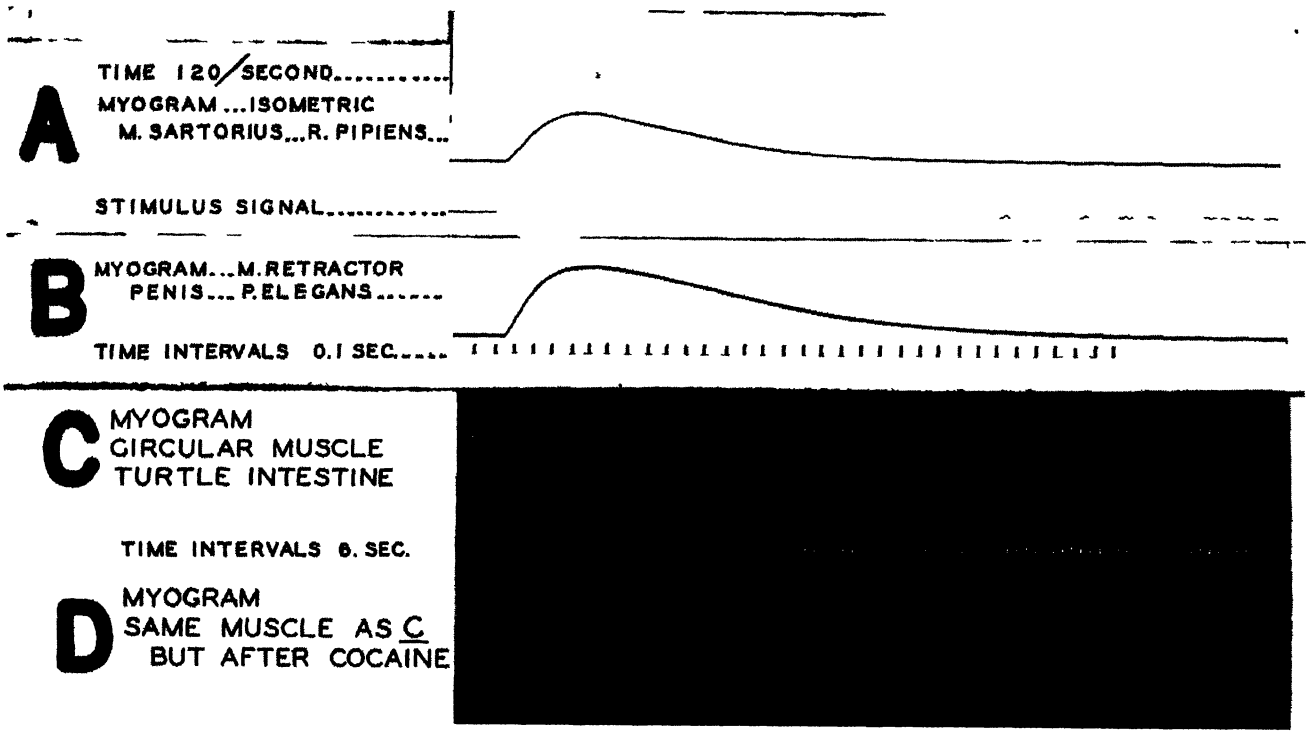


FIGURE 1 Records of various muscles to show similarity of general course of twitch response A and B, isometric recording, C and D, recording by a spring lever allowing moderate shortening

a force F. The forces existing at any time are described by a linear differential equation of the form:

$$m\frac{d^2y}{dt^2} - r\frac{dy}{dt} + n(s + y) + mg - u(b - y) - F = 0, \qquad (1)$$

where m is mass;
r is the overall damping constant for the system;
n is the elastic or "stretch" constant of the lever;
s is the displacement of the lever at rest;
y is the displacement of the end of the muscle;
u is the muscle stretch constant;
b is the strain of the muscle at rest; and
F is the active force developed by the responding muscle.

Now suppose F, the tension developed above resting tension by the muscle at any instant, is proportional to Q, the amount of a substance, B, in the system

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \rightarrow,$$

where a given amount of A is instantaneously released or made available at zero time, and the reactions are apparently of the first order. The following equation then results:

$$m\frac{d^2y}{dt^2} + r\frac{dy}{dt} + y(n + u) = C''[\epsilon^{-k_1 t} - \epsilon^{-k_2 t}]. \qquad (2)$$

Integration yields an approximate solution for plotting of the form:

$$Y = M\epsilon^{-k_1 t} - R\epsilon^{-k_2 t} - (M - R)\epsilon^{-at}, \qquad (3)$$

where

$$M = \frac{C}{(k_1^2 - 2ak_1 + z)} \qquad R = \frac{C}{(k_2^2 - 2ak_2 + z)}$$

$$a = \frac{r}{2m} \qquad C = \frac{C''}{m} \qquad z = \frac{n + u}{m}.$$

The displacement-time course of the muscle would thus be described by an equation containing three exponential terms, in which two of the three time constants are concerned with chemical reactions which act as limiting factors in a chain. For the third exponential term, the time constant is determined by physical conditions of the system, including damping.

An electrical model (FIGURE 2), constructed for convenience in fitting curves, was found to yield wave forms which fitted recorded twitch curves from sartorius muscles as precisely as the method of projection, tracing, and plotting permitted. With such a model, making use of a condenser-resistance network (FIGURE 2A), substitution of R_1C_1 for

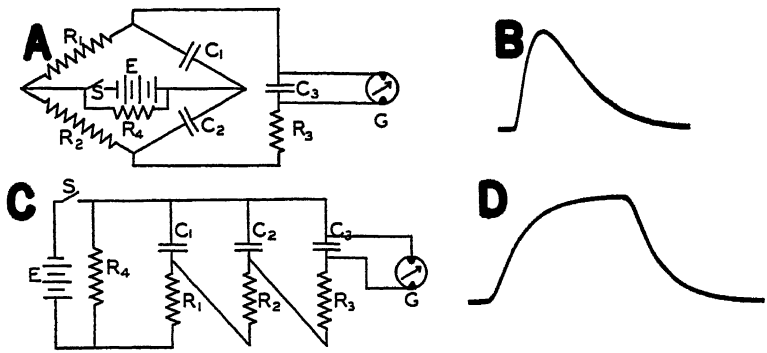


FIGURE 2 A and C, electrical models used with cathode ray oscillograph for recording wave forms corresponding to theoretical equations for twitch and tetanus response, respectively; B and D, photographic records made from models and indicating the theoretical form of twitch and of tetanus responses respectively

1 k_1, of R_2C_2 for $1/k_2$, and of R_3C_3 for $1/a$ yields an equation for a voltage-time curve of the form obtained by plotting from EQUATION 3. A fourth exponential term, which might be expected from the form of the differential equation, was neglected, on the assumption that the system was operating under conditions close to critical damping.

With the lever-muscle system originally used, there appeared to be good agreement between the damping constant of the system determined directly and the value for the third exponential constant found by empirical curve fitting, this being the one mainly manifest in the curves, by its contribution to the foot in the early part of the rise. In later experiments, the forms of the foreloaded and afterloaded twitch curves were compared, since, in the case of the foreloaded response, the form of the foot would depend not only upon exponential terms concerned with viscosity and inertia, but also upon those governed by the position of the tension-producing substance in the chain of chemical reactions. In the case of the afterloaded response, the foot would conceivably involve only the viscosity and inertia factors. For a more general treatment than that previously used, it was shown (Schoepfle and Gilson, 1945) that the balanced forces can be described

by a linear differential equation with constant coefficients and of the form:

$$A_o \frac{d^n y}{dt^n} + A_1 \frac{d^{(n-1)} y}{dt^{(n-1)}} + - - - A_{(n-1)} \frac{dy}{dt} + A_n y = F(t). \qquad (4)$$

This yields a general solution:

$$Y = D(t) + E(t). \qquad (5)$$

The order, n, of EQUATION 4 depends upon the number of series springs and centered masses involved, increasing as the number of springs and masses increases. For the afterloaded response, the function $F(t)$ is replaced by $F_1(t + \tau)$, where τ is the time of pick-up. Hence, Y becomes Y_1 and EQUATION 5 becomes:

$$Y_1 = D_1(t) + E_1(t + \tau). \qquad (6)$$

Solutions of the equations involve sums of exponential terms, since oscillations do not appear. In both instances, foreloaded and afterloaded, exponential time constants derived from the left-hand member of EQUATION 4 will be identical, though the coefficients will vary, depending upon the initial conditions at $t = 0$ and $t = \tau$.

The pronounced difference between semilogarithmic plots for the graphically analyzed foot of the isometric twitch record, in the two cases, shows that $E(t)$ and $E_1(t + \tau)$ must involve more than two exponential terms if there is to be precise description of the foreloaded foot. Hence, $F(t)$ must involve more than two terms. Thus, the substance critical for tension development must occur at a position later than second in the chain of limiting reactions. Conceivably, one or two of the exponential terms in the right-hand member might be concerned with diffusion processes. However, the differences between the time course of the response of a fast, striated and that of a slow, smooth muscle are so tremendous as to throw question upon the importance of diffusion as an important factor in limiting the speed of the response.

Tension changes occurring during the latent period have been examined. The onset of rise of the simple theoretical curve is gradual. Nevertheless, with a variety of recording methods, there elapses a time approaching a minimum of 3 msec. between the application of a short duration shock and the beginning of any recorded tension development.

With a highly sensitive recording system, such as the piezo-electric crystal, the latency relaxation (L. R.), studied by Sandow (1944a), may appear. The cause of its appearance may be attributed to factors somewhat different from those which Sandow has postulated.

The suggestion has been made (Schoepfle and Gilson, 1945) that L. R. is, in a sense, an artifact due to a fluid wave disturbance set up by the mechanical response of the muscle, exerting a force in a direction which may be opposite to that of the tension change of the active muscle response and propagated at a rate greater than that of the spread of muscle activity. This idea is supported by the photographs of Bethe and Happel (1923), and Happel (1926). Evidence for such a possibility follows: First, if the muscle response is reproducible and includes L. R., recorded twitch tension curves, both of L. R. and of the positive tension components, should show constancy of form. Where L. R. is recorded, its presence introduces a variable and inconstant early region in the tension curve, according to the position of the stylus. Second, if a muscle be laid on a glass plate and across it, transversely, there be laid the stylus of the crystal recorder, the stimulating cathode being at another point along the muscle, the record of L. R. begins gradually and attains considerable extent, the positive limb of the curve being delayed for several tenths of a millisecond. If, however, the stylus is made the stimulating cathode, the record shows only a slightly earlier onset of L. R., but much less depth of the negative wave, combined with an earlier and more abrupt onset of the positive phase. These differences are those which should appear if the explanation offered here is correct. Third, L. R. was reproduced in an inorganic model. A light rubber finger-cot was filled with water and suspended by its upper extremity from the stylus of the crystal recorder. To the lower extremity there was attached a metal bar. If the bar was struck with a rubber hammer, the record obtained closely resembled the L. R. found for muscle. If a metal rod was used as a hammer, giving a more sudden impact, the record showed more or less tendency to oscillate during the L. R. The overall viscosity of the water-filled finger-cot is less than that of muscle, and the tendency to oscillate is greater. Otherwise, the L. R. appears similar and to have similar origin, namely, the fluid wave disturbance set up by the sudden onset of a more or less local mechanical disturbance. Even the use of massive electrodes in the attempt to stimulate the entire length of the muscle instantaneously, as Sandow has done, fails to make the situation clear. Here again, one faces as a fact the beginning of some mechanical disturbance which could set up a fluid pressure wave. The sudden stretching of a water-filled finger-cot presents no obvious local disturbance, yet a complicated preliminary displacement of the tension record is found and may be attributed to a small, localized disturbance.

EFFECT OF CHANGED CONDITIONS UPON THE FORM OF THE TWITCH CURVE

Effect of Initial Tension

Increase of initial tension has been reported to cause a delay of fall from peak tension (Hartree and Hill, 1921). We have obtained records in which a slight effect of this sort is found. On the other hand, records have been made with carefully controlled technique, especially in summer, in which large increase of resting tension has resulted in no change in the form of the recorded curves, if these were plotted to give constant height above resting level. This has been the usual result in our experiments.

Effect of Fatigue

Repeated stimulation of excised frog muscle characteristically causes a marked slowing of the twitch response and a diminution of peak tension. However, with muscles in good condition, and again especially in summer, there may be a pronounced fall of twitch tension in the course of a continued series of twitches, with no slowing of the twitch response.

Effect of Shortening

Since, for small displacements, tension developed can be considered as independent of displacement, this approximation was employed in the development of EQUATION 2. The effect of significant shortening was investigated by use of a torsion wire provided with a long lever arm, so that the muscle could be attached close to, or more remote from, the axis. Thus, if a constant tension were developed by the muscle, the use of the longer lever arm would give greater torque and correspondingly greater shortening of the muscle than when a short lever arm was employed. The effective stretch constant of the lever (n, grams tension per mm. shortening) would, of course, be changed. Such experiments have not resulted in the finding of significant differences between the replotted curves of muscles permitted various degrees of shortening. It would be expected that a slight change in form of the response curve would be introduced by the change of the "stretch constant" for the lever, but this would be slight, and sufficiently precise reproducibility has not been attained to make it certain that such changes were clearly demonstrable.

Effect of Iodoacetate

It seemed probable that iodoacetate, acting as an inhibitor in the transfer of phosphate, might slow the course of the twitch response.

No such effect was found. Sandow (1944b) has reported that he finds no consistent variation of critical times in records from rested muscle soaked in 1:40,000 solutions of iodoacetic acid. Several possible conclusions could be drawn from such results: First, that the iodoacetate produces its contracture by causing shortening of the connective tissue of the musle. Second, that the action of the iodoacetate is all-or-none in nature, so that its action on a given fiber is to produce maximal contracture, with no effect on fibers not yet penetrated, this remainder then contracting as normally. Third, that, as contracture develops in a fiber, a part of the contractile substance is removed from availability for twitch response, although the time course of the twitch yielded by the remainder is unchanged. Fourth, that iodoacetate inhibits reactions which may mediately, but not immediately, be concerned with the chemistry of tension development. The last two possibilities appear to be the more probable at the moment. Recently, Price and Schoepfle (personal communication) have found that arsenate and fluoride, also inhibitors for the ATP system, produce no change in the time course of the twitch of the turtle retractor penis muscle.

FORM OF THE TETANUS CURVE

If a muscle be stimulated repeatedly and within a somewhat limited frequency range, one obtains the characteristic tetanic response. The tension record rises to a level which, for frog muscle, is about twice that for the single twitch. A plateau tension level is reached, and this is maintained as long as stimulation continues, or until fatigue causes a loss of tension. It is convenient, here, to consider that the plateau tension remains constant. By a theoretical treatment similar to that used for the twitch, a mathematical formulation can be obtained to describe the form of the tetanus curve to be expected if the critical substance which determines tension is supplied, at a constant rate, over a period long enough to permit attainment of steady state. The equation describing the rise of tetanus tension when stimulation begins has the form:

$$Y = D(R\epsilon^{-k_2 t} - M\epsilon^{-k_1 t} + N\epsilon^{-at}), \tag{7}$$

where $R = \dfrac{k_1}{mk_2(k_2^2 - 2ak_2 + z)}$ $\qquad M = \dfrac{k_2}{m(k_1^2 - 2ak_1 + z)}$

$$N = \frac{k_1 - k_2}{k_2(u + n)} - \frac{k_1}{mk_2(k_2^2 - 2ak_2 + z)} + \frac{1}{m(k_1^2 - 2ak_1 + z)}$$

$$D = \frac{N_1}{(k_1 - k_2)},$$

and N_1 depends upon the amount of the initial material in the chemical chain made available per second. The theoretical curve for the tension fall follows an inverse curve, so that records can easily be obtained by use of an electrical model (FIGURE 2D).

Using data obtained from twitches of a frog sartorius preceding or following a brief tetanus, it has not been possible to find theoretical curves which give good agreement with either the rising or the falling phase of tetanus records actually obtained. Because of irregularities which seemed to be present in the tetanus curves analyzed by Hartree and Hill (1921), an attempt was made to investigate the mechanical changes leading to establishment of the tetanus plateau. By an electronic control device, muscles were stimulated at various frequencies and by one, two, three, etc., shocks, respectively. Records from muscles stimulated at two frequencies, 25 per second and 100 per second, may be considered here. In both cases, the succession of stimuli was sufficiently rapid to give marked summation of successive mechanical responses. The records were photographed, projected, and traced. The tension contribution by each shock was then determined by graphical subtraction. Using the simple theoretical equation for rise of tetanus tension, values were plotted, and yielded equal tension-time increments for each unit period of time. The actual experimental records showed a first response which was large. Depending upon the frequency of stimulation and the condition of the muscle, the second, and even the third, responses might be large. Thereafter, the increments of response remained of fairly equal size, but were much smaller than the initial response increment and, in frog muscle, showed a progressive slowing. The actually recorded course of tension rise deviated from the simple theoretical curve in a way to be expected from the fact that the muscle does not establish its full tetanus tension by continued liberation of response increments as great as those of the initial one or two responses.

The deviation from the theoretical curve of the actual course of fall from the tetanus plateau, when stimulation ceases, also appears largely due to the fact that the tension responses at any frequency of stimulation are incremental and not continuous in nature, and to the changes in the course of the mechanical responses which took place as the tetanus continued, the fall from tetanus beginning less promptly and proceeding more slowly than would be expected from the measurement of a twitch elicited shortly before or after the period of tetanus. The more closely the twitch follows the tetanus, the better is the agreement. In the rat (*M. triceps surae*), the fall from tetanus may

be more rapid than would be predicted from the form of the simple twitch, a finding which may be in keeping with the facts that such muscles show a marked potentiation (see Brown and von Euler, 1938) and a quickening of the response as the result of a brief tetanus. This potentiation has occasionally been observed in frog sartorius preparations in oxygenated, well-buffered saline solution.

CHANGE OF THE TIME COURSE OF THE TWITCH RESPONSE DURING THE DEVELOPMENT OF FATIGUE, ETC.

In the preceding discussion of the twitch response, emphasis was placed on the stability of the time course of the response, which has been present under a variety of experimentally changed conditions. Even for the sartorius of the frog, such constancy frequently is lacking. For early responses, the curves may show a residual contracture which introduces a marked deviation from the simple theoretical curve. Tiegel (1876) observed that this phenomenon appeared particularly during the latter part of February and through the entire month of March. We have obtained a striking example of this in records which were taken from an experiment performed November 25th. Fatigue, especially in winter animals, usually causes a progressive spreading-out of the response curves. Certain poisons, among them caffeine, may cause a great change in the time course of the response.

The change in time course of the response of the frog gastrocnemius muscle serves to illustrate the very great change of twitch form which may occur. The first response after the muscle is prepared may be a quick response, the form of which agrees well with that of the theoretical curve. Thereafter, a progressive spreading-out of the curve develops (PLATE 1). This is particularly apparent in the slowing of the falling phase. Parkinson (1933) concluded that "a highly rested condition is unfavorable" to the development of the spreading of the curve. Between the crest and that part of the relaxation phase which shows an approximately exponential fall, there is a sloping plateau terminating in a definite shoulder. The shoulder is reminiscent of the "angle" of Fulton (1925). Cooper and Eccles (1930) have pointed out that the form of the curves recorded by Fulton was distorted, because of friction in the bearing of the torsion lever which he used. Nevertheless, it seems probable that the "angle" in many of Fulton's published records is the shoulder referred to here, but exaggerated by the lever friction. A sign of the shoulder can often be found very early in the course of a fatigue series. A twitch curve of this sort is quite different from that postulated as representing the simple theoretical

response. A curve of the type described here can be developed from the theory, by supposing that formation of precursor substance critical for contraction does not take place instantaneously, but that the material is continuously made available, although in somewhat diminishing quantities, over a rather extended period. In other words, the tension response follows a course resembling that of a brief tetanus complicated by fatigue. Action potentials recorded from such muscles show only a single recorded spike potential, at amplifications considerably above that necessary to record the synchronized spike. It appears that there occurs a warming-up process, resulting from activity of the muscle and bringing about a great prolongation of the period of sustained tension. If the muscle is in poor condition or badly fatigued, the shoulder is lost or masked by the slowing-down of the response. It was apparently from a muscle of this type (*M. semimembranosus* of the frog) that Fick (1882) recorded the responses used to illustrate his conclusion that, in isometric contraction, the development of tension is faster and the crest attained more quickly than is true for the isotonic contraction. The forms of his records are, in fact, quite as should be expected, if one considers inertia and acceleration. The important point brought out by such records is that active relaxation begins at about the same time for the isometric and the isotonic records. Although such spreading of the response, as apart from mere slowing, seems due to prolongation of tension in a single response of frog muscle, mammalian muscles may show a similar form of "twitch," due to repetitive stimulation (of the nerve) by a single shock.

DISCUSSION

As long as our experiments were limited to "isometric" responses, at approximately constant muscle length, the theory presented has seemed to be in agreement with the experimental facts observed. Without modification, the theory has not provided a description of the tension changes which occur when there is a pronounced change of muscle length during contraction. Consideration of the distribution of the forces acting and reference to the simple stress-strain diagram for muscle show that such provision could be made. In the material to follow, assumption will be made of the validity of the conclusions of Banus and Zetlin (1938), and of Ramsey and Street (1940), that the tension contribution of resting muscle is by sarcolemma, connective tissue, etc. For active muscle, either at the same or at a shorter length, if connective tissue and active muscle substance act in parallel,

the tension must be determined by the summed contributions of connective tissue and of active muscle substance.

The stress-strain relationships for muscle are diagrammatically represented in FIGURE 3 as a straight line figure. At first sight, the diagram appears to be the plot for two elastic or spring systems, one system corresponding to the muscle in its resting state, the other to the muscle in its active state This was pointed out, a hundred years ago, by Weber (1846). Such a diagram implies that the active muscle is less elastic, actually a weaker though shorter spring than the resting muscle, so that it shows greater lengthening per unit increase of tension. Thus, the added tension contribution by the active muscle substance alone is not that of an added parallel spring, but of a material or force such that there is plotted a stress-strain curve (line MAN) with a negative slope, rather than with the positive slope to be found for a true elastic substance. Moreover, the active muscle substance either is material having no resistance to stretch at the length determined by the point of crossing of the two lines (X) or at this length

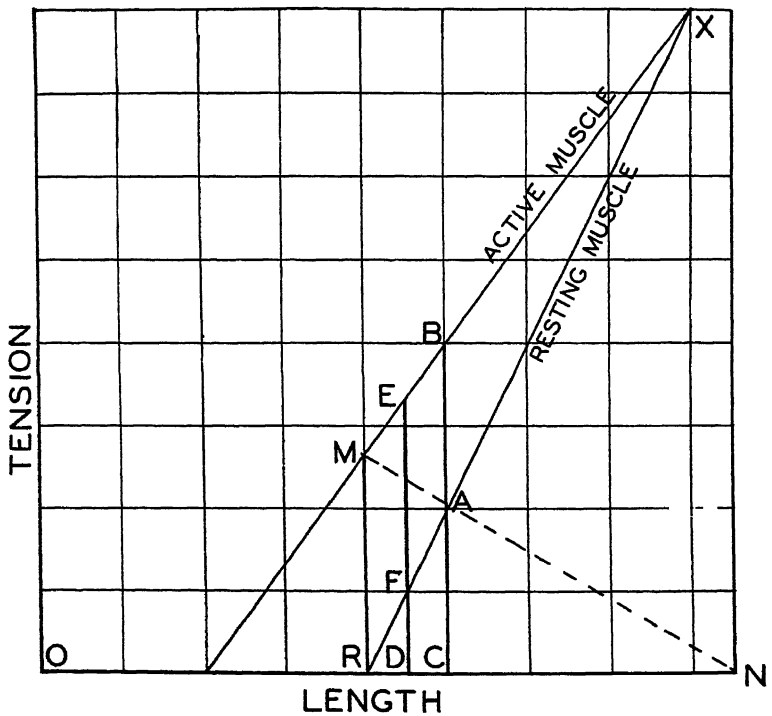


FIGURE 3. Diagrammatic tension-length diagram for muscle, drawn as a straight line figure.

no effective contractile substance is formed, or the diagram does not furnish a true description of the elastic conditions. The last situation seems to offer at least a partial answer, since the elasticity of active muscle can be shown to be the same or nearly the same as that of resting muscle at the same tension (Schoepfle and Gilson, 1946). This is difficult to understand in terms of a concept of simple axial shortening of longitudinal elastic structures which express their tension by force applied at the ends of the muscle fibers

When the muscle responds under experimental conditions which permit shortening, the tension developed above resting tension is less than that found under isometric conditions. From FIGURE 3, it is seen that, if a stretched muscle is stimulated and in responding shortens, its final tension must be less than that which would have been attained with the same resting length, but with the muscle responding under isometric conditions. Under this shortened condition, however, the tension contribution by connective tissue, etc., is less than would have appeared under the isometric condition. Thus, if the active contractile material is exerting a simple axial tension in parallel with the connective tissue structures, it must develop, when shortening, not a lesser, but a greater effective tension contribution than when response occurs under isometric conditions On the other hand, if the primary twitch force were exerted laterally and thus effectively in series with the connective tissue, a quite different situation would appear (Schoepfle and Gilson, 1946). If the stress-strain diagram for resting muscle were correctly drawn as a straight line diagram (FIGURE 3), F, the effective active force of the muscle would diminish to reach zero at the intersection of the lines for resting and for active muscle, respectively. If the stress-strain diagram for resting muscle is correctly drawn as a curved line, concave upward, and for the active muscle as a curved line, concave downward (see Bard's text, p. 30, or Hill, 1913b), the detailed relationships would be quite different. If the curves had a still more complicated form, F might even be constant at any length of the muscle. Further quantitative data on this subject are needed.

CONCLUSION

The material presented above has offered a detailed description of the course of muscle tension during a response. The theoretical treatment, largely empirical though it be, has offered a background against which it has been possible to examine records from a variety of muscles

under a variety of conditions. These studies have emphasized the constancy of form of the twitch response which may persist, despite changes in state of fatigue, of resting tension, and of shortening permitted during the response. They have also emphasized the general similarity of the course of the simple or basic response of a variety of muscles, and they have shown that one can recognize as more or less separable variables: (1) the tension developed; (2) the quickness of the simple contraction and relaxation process; and (3) the duration of the response, in a sense essentially that formerly used by Fulton. Residual contractures may, perhaps, be differentiated from mere prolongation of the response proper. Distinction must be made between those responses which are complicated by processes which arise within the muscle as tissue, and those arising within the muscle as organ. The retractor penis of the turtle, when freshly prepared, usually shows a delayed relaxation which gives a twitch response curve quite different from theoretical. The phenomenon may be due to a Tiegel contracture, but it persists through many more responses than does the Tiegel contracture in frog muscle. Such a muscle soaked in Ringer solution overnight loses its residual contracture following a twitch, and yields a response curve which closely approaches the theoretical curve form. The smooth circular muscle of the turtle intestine, when freshly prepared, may show a prolongation of contraction similar to that described for the frog gastrocnemius (FIGURE 1C). After treatment with cocaine, the muscle responds to a single stimulus with a twitch which is much slower than that for skeletal muscle, but yields a curve form close to that expected under the theory (FIGURE 1D). The change of response, in this case, may be due to the action of the cocaine on nervous structures, rather than to an action on the muscle proper.

No formal attempt has been made to align the theoretical description outlined above with that of Hill (1938), or to introduce Hill's dynamic constants into our analysis. It is not impossible that this could be done. The final equation developed from such treatment would certainly be of a form less simple than either Hill's equation or EQUATION 7 of this paper. Although the Hill equation does not yield a close fit to the curves of rising tension in tetanus, his treatment did not give consideration to the fact that, whatever the rate of stimulation in tetanus, the curve of rising tension is determined by more or less complete fusion of a series of responses to the successive stimuli. The first mechanical response in a tetanus is from resting muscle, and

it exceeds in magnitude those responses which follow later. Under the theory, one might expect the tetanic response to show diminution in metabolic turnover *per stimulus*, as the frequency is increased, and a fairly constant turnover *per second*, regardless of frequency. That this, in fact, is possible, is indicated by the observation of Hill (1913a) of an approximate constancy of maintenance heat during tetanus at varying frequencies of stimulation. Bronk (1930) more recently showed an increase in efficiency at higher rates of stimulation. Bronk worked, however, in the low frequency range, so that his results are also consistent with the above postulate. For the case of the twitch, the tension-time area, as well as the peak-developed tension, would be determined by the effective time constants of the limiting reactions, as well as by the amount of substance mobilized, and so would become a less direct measure of the chemical cost of contraction. Maintenance of tetanus plateau tension is, of course, to be thought of as a dynamic equilibrium, with the tension remaining constant at the cost of continued metabolic activity.

It does seem probable: (1) that the twitch response is determined by the amount available of a critical substance which is formed in a series of chain reactions; (2) that the time course of the response, and incidentally the tension developed, depend upon the velocity constants of at least two (and probably more) limiting steps in such a chain; (3) that, in tetanus, the limiting factor for maintenance of tension is the amount of material per second which is made available to this chain series, and that, following the first very few responses, the amount transferred per stimulus is much less than that made available for the resting twitch; (4) that the stress-strain diagram for resting and active muscle describes overall performance, but does not furnish a true diagram of elastic conditions in the two states. Since sudden stretch or shortening, during a twitch, indicates little or no change of elasticity *at a given tension* and little, if any, change in viscosity of the muscle (frog sartorius and turtle retractor penis), one finds serious objections to the concept that muscle contraction is due to increased tension of fibrillar material in simple parallel arrangement with sarcolemma and connective tissue structures. The need for some alternative theory to explain the development of axial tension is indicated. Moreover, the simple theory outlined above must be modified to include consideration of the changes of axial tensions developed at different resting lengths of the muscle.

Many conclusions explicit in this material and implicit from it must,

to have significance, be in accordance with pertinent data obtained in physical and chemical experimentation. The future value of this analysis must, therefore, depend upon the extent to which it finds accord with such experimental material.

LITERATURE CITED

Banus, M. G., & A. M. Zetlin
1938. The relation of isometric tension to length in skeletal muscle. J. Cell. & Comp. Physiol. **12**: 403–420.

Bard, P.
1941. Macleod's Physiology in Modern Medicine: p. 30. 9th ed. C. V. Mosby Co. St. Louis.

Bethe, A., & P. Happel
1923. Die Zerlegung der Muskelzuckung in Teilfunktionen I. Pflüg. Arch. **201**: 157–181.

Bronk, D. W.
1930. The energy expended in maintaining a muscular contraction. J. Physiol. **69**: 306–315.

Brown, G. L., & U. S. von Euler
1938. The after-effects of a tetanus on mammalian muscle. J. Physiol. **93**: 39–60.

Cooper, S., & J. C. Eccles
1930. The isometric responses of mammalian muscles. J. Physiol. **69**: 377–385.

Fick, A.
1882. Mechanische Arbeit und Wärmeentwicklung bei der Muskelthätigkeit. F. A. Brockhaus. Leipzig.

Fulton, J. F.
1925. Some observations upon the electrical responses and shape of the isometric twitch of skeletal muscle (intact). Proc. Roy. Soc. London B **97**: 424–431.

Gilson, A. S., Jr., S. M. Walker, & G. M. Schoepfle
1944. The forms of the isometric twitch and isometric tetanus curves recorded from the frog's sartorius muscle. J. Cell. & Comp. Physiol. **23**: 119–130.

Happel, P.
1926. Die Zerlegung der Muskelzuckung in Teilfunktionen II. Pflüg. Arch. **213**: 336–351.

Hartree, W., & A. V. Hill
1921. The nature of the isometric twitch. J. Physiol. **55**: 389–411.

Hill, A. V.
1913a. The effects of frequency of excitation upon the relation between mechanical and thermal response in muscle. J. Physiol. **46**: vii.
1913b. The absolute mechanical efficiency of the contraction of isolated muscle. J. Physiol. **46**: 435–469.
1938. The heat of shortening and the dynamic constants of muscle. Proc. Roy. Soc. London B **126**: 136–195.

Parkinson, J. L.
1933. The effect of activity on the form of the muscle twitch. J. Physiol. **78**: 106–112.

Ramsey, R. W., & S. F. Street
1940. The isometric length-tension diagram of isolated skeletal muscle fibers. J. Cell. & Comp. Physiol. **15**: 11–34.

Sandow, A.

1944a. Studies on the latent period of muscular contraction. J. Cell. & Comp. Physiol. **24**: 221–256.

1944b. Study of the effect of pH, tissue poisons, and isotonicity, on the mechanical events of the latent, contraction, and relaxation periods of skeletal muscle contraction. Year Book, Am. Phil. Soc.: 182–185.

Schoepfle, G. M., & A. S. Gilson, Jr.

1945. Some observations on the configuration of the free loaded and after loaded muscle twitch. J. Cell. & Comp. Physiol. **26**: 119–130.

1946. Elasticity of muscle in relation to actively developed twitch tension. J. Cell. & Comp. Physiol. **27**: 105–114.

Tiegel, T.

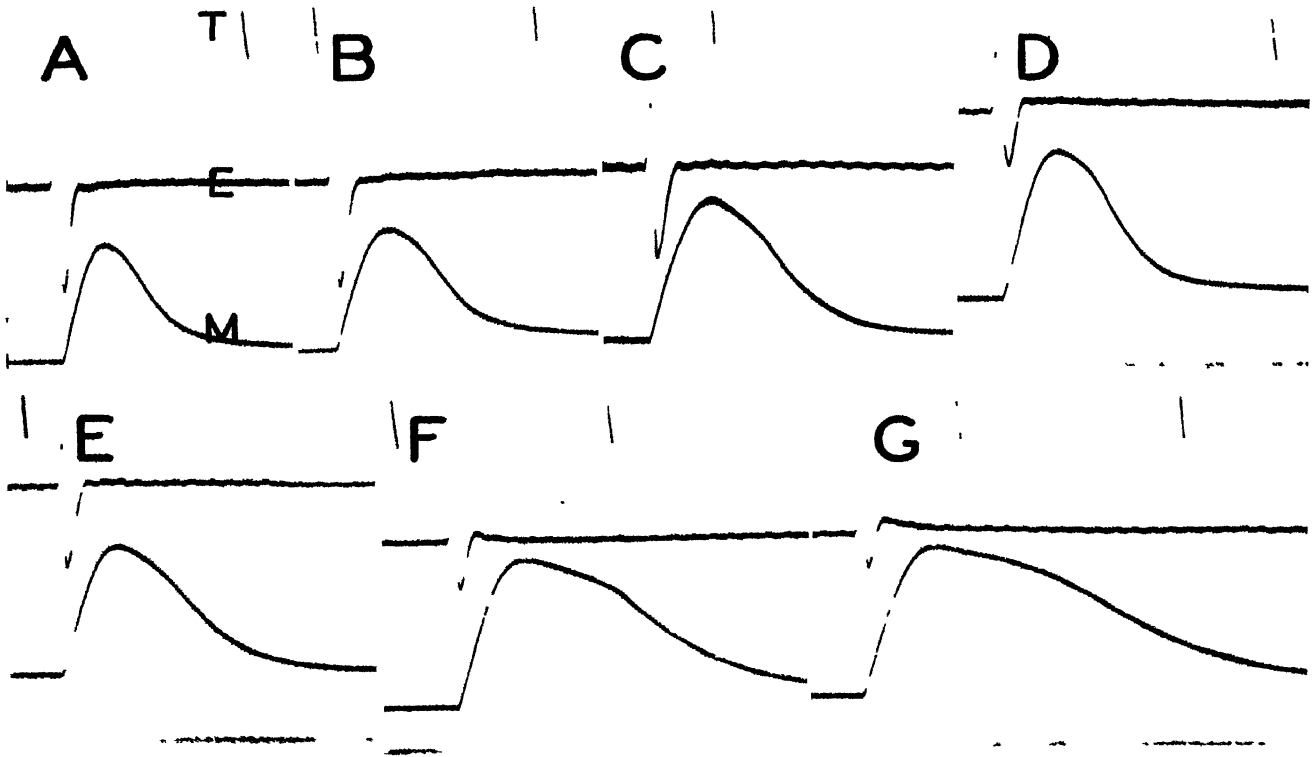
1876. Ueber Muskelcontractur im Gegensatz zur Contraction. Pflüg. Arch. **13**: 72–83.

Weber, E.

1846. Muskelbewegung. Wagner's Handwörterbuch d. Physiol. **3**(2): 115. Fr. Viehweg & Sohn. Braunschweig.

PLATE 1

Isometric myograms and electrograms from gastrocnemius muscle of frog. Indirect stimulation. Nerve was stimulated at intervals, irregularly, to give 105 responses, the first, last, and five intermediate responses having been selected for reproduction. Records show significant change of mechanical response with but a single apparent spike potential in each case. T, time intervals of 1/5 sec.; E, electrogram recorded by needle electrodes thrust into tendon and belly of muscle; M, isometric mechanogram.



GILSON ET AL: TIME COURSE OF TENSION DEVELOPMENT

MUSCULAR CONTRACTION AND RUBBERLIKE ELASTICITY

By Eugene Guth

Polymer Physics Laboratory, University of Notre Dame, Notre Dame, Indiana

INTRODUCTION

The problem of rubberlike elasticity is deeply interconnected with all those sciences which make a fundamental inquiry into organic life. These sciences are biochemistry, biophysics, physiology, pathology, and genetics, and they deal with nerves, muscles, vitamins, proteins, and chromosomes.

Rubber belongs to a class of substances which nature uses as a framework for all living things. The similarity between rubber and resting muscle has long been recognized. It is well brought out by the fact that resting muscle develops heat when stretched rapidly, just as rubber does. Of course, the behavior of muscle is, in general, much more complicated than that of rubber. Nevertheless, in certain essential aspects, rubber may be considered as a simple prototype or model of muscle.

From this point of view, the study of rubber may be considered to be an important link in the monumental attempt of the human mind to unveil the mystery of organic life. In the tree, *Hevea brasiliensis*, nature is the manufacturer of rubber. After the discovery of rubber latex and its transformation into solid crude rubber by the colloidal process of coagulation, the next important development in rubber technology was the discovery of vulcanization by Goodyear. Though unknown to Goodyear, vulcanization is an imitation of an artifice of nature to preserve an organic material from decay and make it more enduring. Wool protein or keratin contains sulfur as a protective agent, just as vulcanized rubber does. Consequently, it may last a long time after the "crude" (*i.e.*, unprotected) proteins of an animal body have disintegrated. The synthesis of rubberlike materials is a successful attempt of man to replace nature as a manufacturer of rubber or to find substitutes for this material.

It is of interest to point out that one of the earliest investigators of rubber elasticity was a physiologist (Schmulewitsch, 1870). In recent years, the quantitative study of muscle has required delicate and accurate instruments. Work of this type has been carried out chiefly by

physiologists with a background in physics and mathematics, like Hill and his school. The physics of rubber was studied many years ago by Joule, Kelvin, Stefan, Boltzmann and Röntgen. Only recently, however, has a satisfactory quantitative theory of the physics of rubber been developed. This theory, developed by James and Guth, gives a quantitative explanation of the long-range elasticity of rubber and of its anomalous thermoelastic behavior. A great deal of quantitative work has been done on the physics of muscle, and only within the last few years has the study of the physics of rubber really forged ahead. This is partly due to the simpler nature of rubber, as compared to muscle, and also partly due to the great impetus which the study of rubber received during the war, as an important part of the war effort. As a result, the physics of rubber may now serve as a guide in the development of the physics of muscle.

The main part of this paper is an account of the present knowledge of the elastic properties of rubber. This knowledge will be applied to certain aspects of the elastic behavior of the resting muscle. Finally, a few comments on certain aspects of the active muscle will be made. No attempt has been made to cover the subject completely. Concerning the behavior of the muscle, references will be made to the other papers presented at the conference. Part A of this article deals with the steady-state elastic behavior and crystallization of rubber. Part B treats, rather briefly, time effects for rubber. Part C deals with the resting, and part D with the active muscle. The treatment in parts A and B is mostly descriptive. Derivations are deferred to the Appendix.

Literature on the elasticity of rubber is given in articles[1, 2, 3] up to about 1942. Literature on the elasticity of muscle is contained in the two recent excellent articles of Ramsey[4] and Fenn.[5] In general, we shall give only references which are not contained in the articles just mentioned.

A. RUBBERLIKE ELASTICITY: STEADY STATE

1. Characteristic Properties of Rubber

The characteristic properties of rubber are: (a) its long range reversible elasticity; and (b) its anomalous thermoelastic behavior. The stress-strain curve for rubber has a peculiar S-shape and hence

[1] **Guth, E.** "Surface Chemistry." Publ. No. **21**, A.A.A.S. 1943.
[2] **Guth, E.** Alexander's "Colloid Chemistry." 113. Reinhold, N. Y. 1944.
[3] **Guth, E., H. M. James, & H. Mark.** "Advances in Colloid Chemistry." 286. Interscience, New York. 1946.
[4] **Ramsey, R. W.** Glasser's "Medical Physics." 783. Year Book Publishers, Chicago. 1944.
[5] **Fenn, W. O.** Höber's "Physical Chemistry of Cells and Tissues." 445. Blakiston, Philadelphia. 1945.

does not follow Hooke's Law. There are three thermoelastic effects for rubber: (1) it generates heat when stretched rapidly; (2) if stretched by a constant load, it contracts visibly when heated; (3) if held at constant length, the stress increases with rising temperature.

FIGURE 1 shows the change in temperature, observed by Joule, when a strip of vulcanized rubber was stretched rapidly. At very low elongations, he observed a slight cooling. At a certain critical extension, defined as the thermoelastic inversion point, the change in temperature was found to be zero. For extensions greater than this critical one, the change in temperature is positive and increases steadily with increasing elongation. In the case of normal solids like glasses, metals, and wood, a cooling effect is observed on fast stretching. By comparison, the thermoelastic behavior of rubber is normal only at small extensions, and it is anomalous above the thermoelastic inversion point.

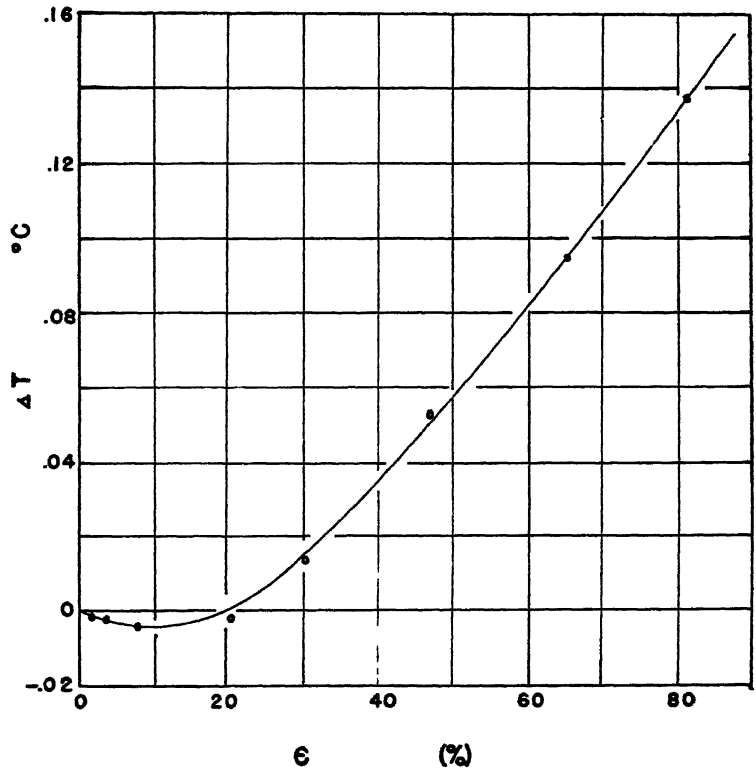


FIGURE 1. Change of temperature ΔT on fast stretching as a function of the strain ε, according to Joule.

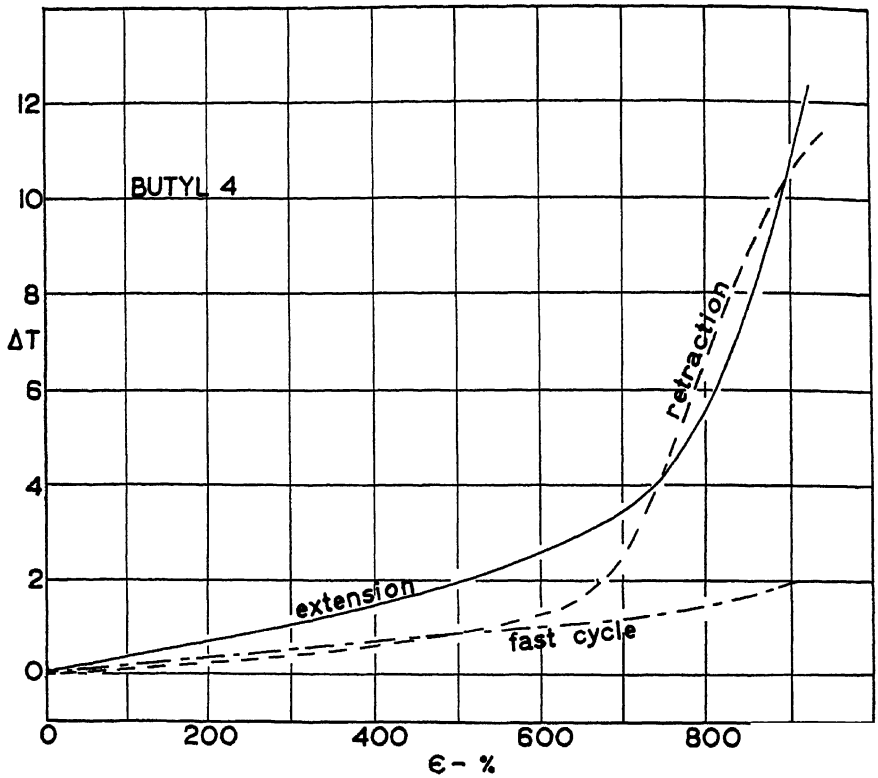


FIGURE 2 Rise in temperature on fast stretching of Butyl rubber.

FIGURE 2 shows the rise in temperature on fast stretching, for Butyl rubber, as a function of extension at large extensions (Dart and Guth[6]). The steep rise at larger extensions indicates progressive crystallization on stretching of the rubber. Besides the rise in temperature on extension, the cooling observed on retraction is also given. This cooling (the negative of which is plotted) was obtained after the rubber was kept at each extension, for which ΔT was measured for one minute. During this time, an additional crystallization took place. This accounts for the hysteresis loop between the extension and the retraction branch. The first crossing of the two branches indicates the start of crystallization, while the second crossing is a sign of a certain saturation. Fast cycle gives ΔT for extension followed by immediate retraction.

[6] Dart, S. L., & E. Guth. J. Chem. Phys. 13: 28. 1945.

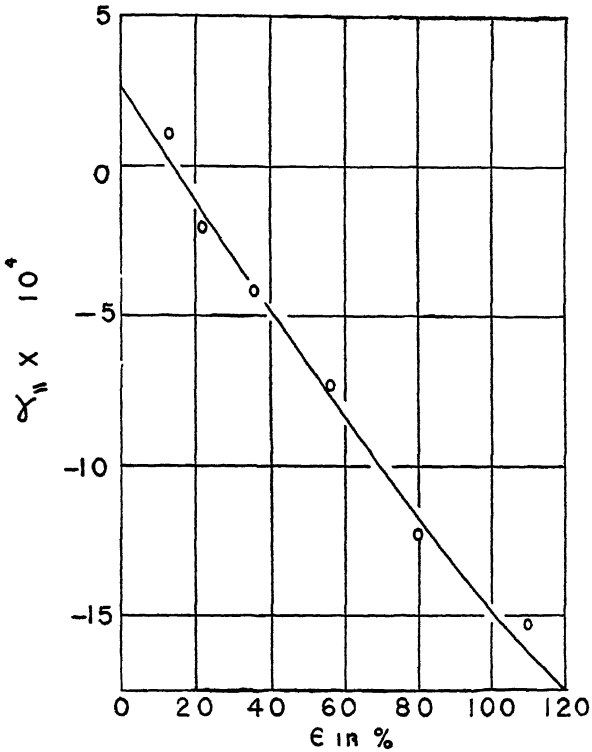


FIGURE 3 Linear thermal expansion coefficient $\alpha_{\parallel}$ parallel to the stress as a function of the strain ε, according to **Joule.**

FIGURE 3 shows the contraction which Joule observed on heating a sample of vulcanized rubber which was stretched by a constant load. Again, a thermoelastic inversion point is observed.

FIGURE 4 shows isometrics obtained by Anthony, Caston, and Guth at Notre Dame. The isometrics exhibit the third thermoelastic effect, namely, the change in stress produced in a strip of rubber when its length is held constant and its temperature is varied.

The first stress-strain curves for rubber were obtained by Villari. FIGURE 5 shows a stress-strain curve obtained by Villari for rubber cord. This curve exhibits nicely the characteristic S-shape of the rubber stress-strain curves. The dashed curve indicates the compression branch.

A proper theory of rubber elasticity must explain quantitatively the three thermoelastic effects and the S-shape of the stress-strain curve. A theory satisfying these conditions was developed by H. M. James

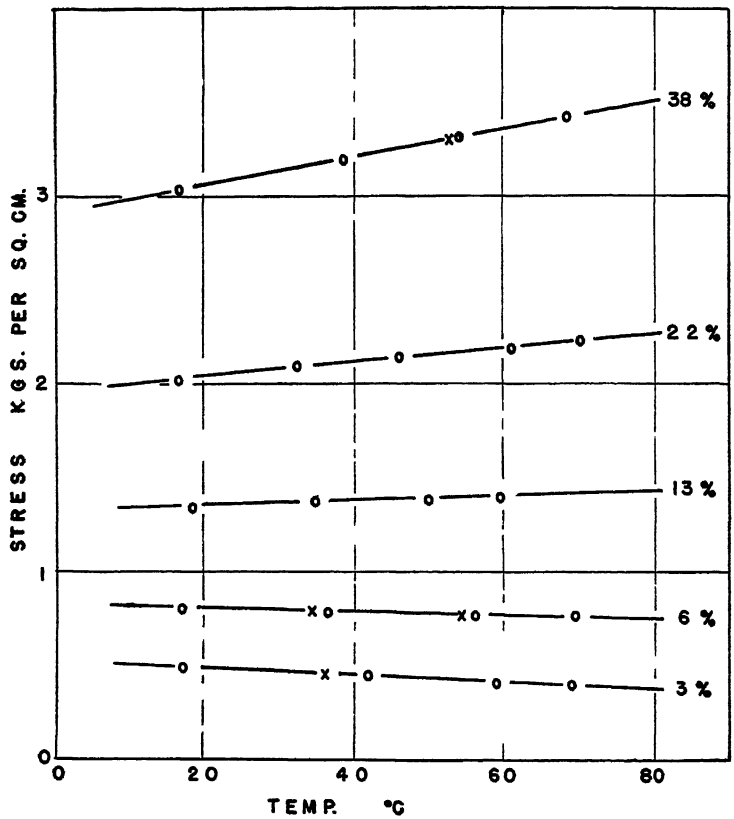


FIGURE 4a Stress-temperature curves (isometrics) at various extensions for an unaccelerated pure gum stock

and the author, and will be described briefly in the later sections of this paper.

2. Resolution of the Stress

The peculiar behavior of rubber shows up very clearly in the resolution of the stress into its additive components, due to internal energy and entropy, respectively. In van der Waals' equation of a real gas,

$$p = -a/V^2 + RT/V - b = p(U) + p(S), \tag{1}$$

the first term $p(U) = -a/V^2$ is due to the internal energy U, and the second term $p(S)$ is due to the entropy S. Similarly, the stress Z may be resolved as follows:

$$Z = Z(U) + Z(S). \tag{2}$$

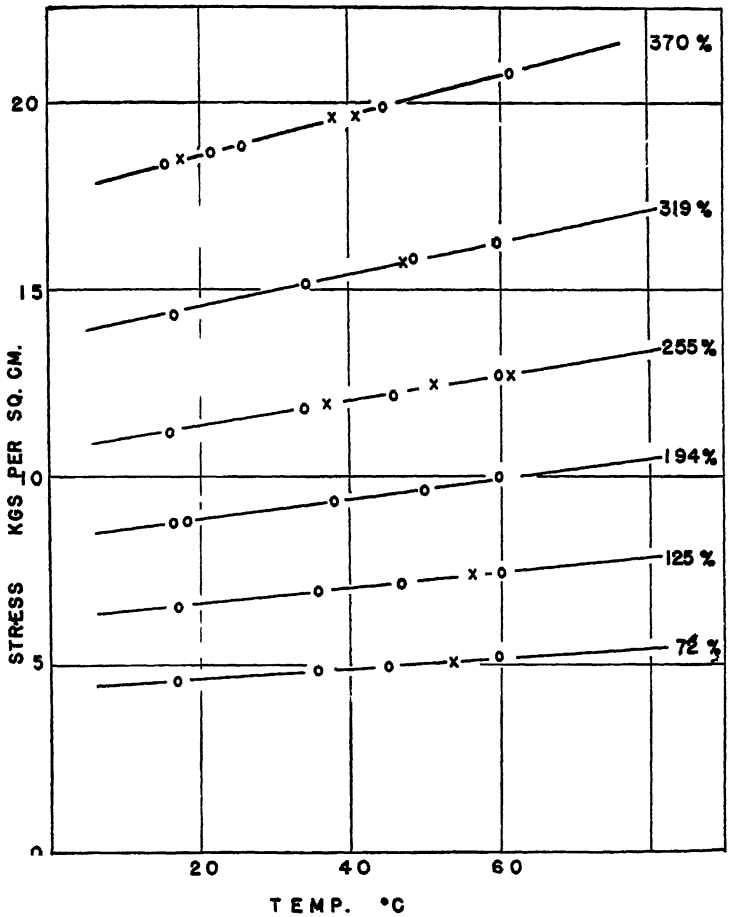


FIGURE 4b Same as FIGURE 4a for larger extensions

Such a resolution may be obtained from observed isometrics using thermodynamics only, *i.e.*, without any knowledge of the structure of the particular material under consideration. Roughly speaking, $Z(U)$ is connected with intermolecular forces, while $Z(S)$ is due to heat motion* (cf. the above identification of $p(U)$ and $p(S)$ for a gas).

A gas may be charactcrized by the fact that $p(S)$ (heat motion) determines the pressure, while $p(U)$ (intermolecular forces) is merely a correction, $p(U)$ becoming, however, increasingly important when the

* More accurately, $Z(U)$ also may depend to some extent upon the heat motion, and $Z(S)$ may also depend upon the intermolecular force (cf. the occurrence of "b" in $p(S)$).

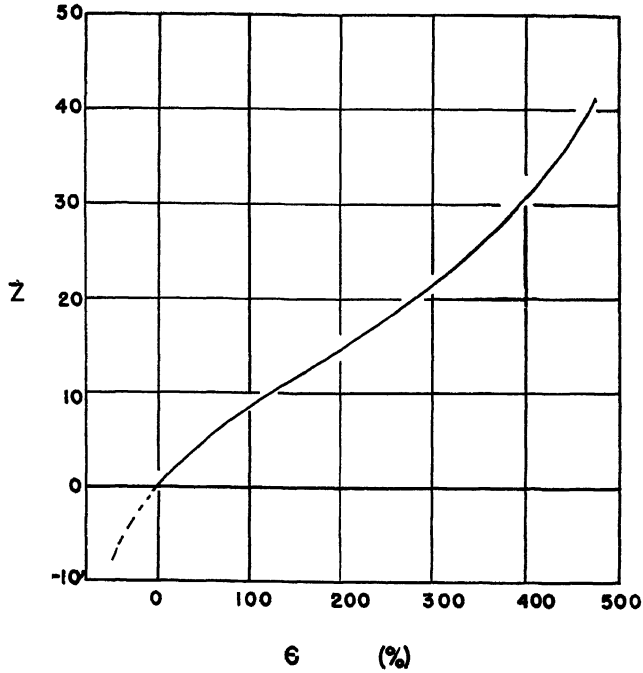


FIGURE 5 Stress-strain curve for a rubber cord, according to **Villari.** The dashed curve represents compression data by **Sheppard & Clapson.**

gas is condensed, *i.e.* when the volume V gets small. This is the more true when the gas undergoes liquefaction. It is seen that the relative contribution of $p(S)$ or $p(U)$ to the total pressure depends upon the volume V (and the temperature T).

In a similar way, we can infer the relative importance of $Z(U)$ and $Z(S)$ for the case of *solids,* once the stress is resolved into these two components. For normal solids, $Z(U)$ is at least of the same importance as $Z(S)$. For rubber, the situation is different. FIGURE 6 shows a resolution for natural rubber, using data obtained by Wiegand and Snyder. For comparison with muscles, one may distinguish three regions. In the first region (0–350 %), $Z(U)$ is practically constant, while $Z(S)$ increases and becomes predominant in the total stress. In the second region, $Z(U)$ becomes strongly negative, while $Z(S)$ still increases. In the third region (550–750 %), $Z(U)$ turns sharply upwards, while $Z(S)$ turns downwards. This is the region where the total stress increases strongly with further extension. The per cent extensions which separate the three regions depend, of course,

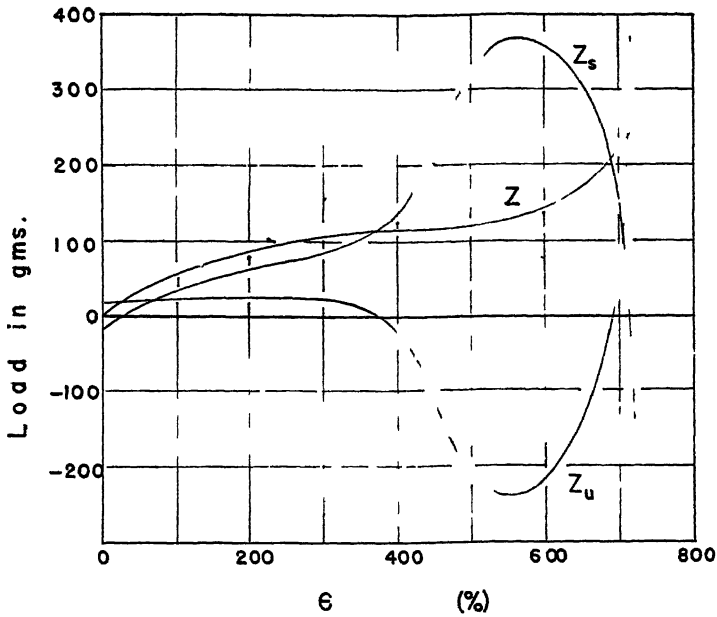


FIGURE 6 Resolution of the stress Z for natural rubber into its two components $Z(S)$ and $Z(U)$.

upon the particular compound used. The existence of a division into the three regions may, however, be more general.*

In the first region, rubber shows an elasticity similar to that of a gas (neglecting, for the moment, the thermoelastic inversion): $Z(S)$ predominates, *i.e.*, in this region, the stress must be mainly due to heat motion of constituents of the rubber. Naturally, one is led to consider the heat motion connected with free rotation in the long rubber chain molecules. There is, however, a non-negligible contribution by $Z(U)$, too.

In the second region, rubber starts to crystallize. With increasing stretch, the amount of crystalline material increases at the cost of the amorphous material. $Z(U)$ is of the same order as $Z(S)$ in this region, as one should expect off-hand.

In the third region, the crystallization process continues up to a certain saturation, at which point, however, there is still some amorphous material left. $Z(U)$ increases rapidly in this region. $Z(S)$ decreases rapidly, *i.e.*, the stress is caused more and more by inter-molec-

* **P. E. Wack** checked at Notre Dame the upward turn of $Z(U)$ after it passed a minimum, though he did not stretch the sample far enough to check the crossing of the zero axis by $Z(U)$ and $Z(S)$. We wish to point out that **Wood, L. A., & F. L. Roth** (J. Appl. Phys. **15**: 781. 1944) did not observe a minimum of $Z(U)$. This matter needs further clarification.

ular forces and less and less by heat motion. We shall see later that resting muscle corresponds to rubber in this third region.

The resolution of the stress permits comparison of the experiment with theory, both for $Z(S)$ and $Z(U)$ separately. The theory of James and the author explains $Z(S)$ and part of $Z(U)$ in the first region fairly adequately. There is no quantitative theory of the behavior of rubber in the second and third regions, which are of major interest in connection with muscles.

3. The Structure of Rubber

There are six main questions connected with the structure of rubber. The first concerns its elementary formula. This may be deducted from purely chemical facts and turns out to be $(C_5H_8)_n$. (The isoprene unit, C_5H_8, was discovered by Faraday.) The second question pertains to the structure within the isoprene groups. Chemistry shows that this structure is:

$$-CH_2-CH=C(CH_3)-CH_2-$$

containing one double bond and three single bonds. The third question concerns the relation of the isoprene units to one another; *i.e.*, whether rubber is in the *cis*- or *trans*-configuration. The fourth question relates to the degree of polymerization, n. Physico-chemical data, especially x-ray analysis, show that rubber is in the *cis*-configuration, and that n is very large. The rubber molecule consists of a very long chain of isoprene units. The fifth question is related to the flexibility of the rubber molecule. In an isoprene group, three out of the four C-C bonds are single, and only one is a double bond. Both experiment and theory show that there can be free rotation around a single C-C bond. On the other hand, a double bond is rigid and defines a certain plane which is characteristic of the group containing the double bond. The term, *free rotation,* as employed here, means only that there are different configurations (stable arrangements) of equal relative probability. These configurations may be separated by energy barriers, so that a free rotation, in the literal sense, does not take place. However, if the barriers are not too high, a relative equilibrium of the different configurations may occur in a short time. The free rotation is simply the result of the heat motion. It will give rise to the persistence of an internal mobility or Brownian motion in solid rubber. *The great length and the great flexibility of the rubber molecules are the essential elements of rubberlike elasticity.*

The above five questions are the ones which arise in connection with the study of more or less dilute solutions of rubber. The

sixth question pertains to the structure of bulk rubber, and it will be discussed in the following sections. Two models will be treated quantitatively. These are: (a) the bundle model; and (b) the network model. The network model is a close approximation to reality. The bundle model is useful for a simple explanation of some essential features of rubberlike elasticity.

4. Bundle Model

In the first semi-quantitative treatment of rubber elasticity by Guth and co-workers, in 1934, a simplified model was used for bulk rubber. This model consists of a bundle of parallel chains or strings running through the whole piece of rubber. The piece of rubber selected is a unit cube whose edges are parallel to the three coordinate axes, x, y, z. If M designates the number of chains per unit cross-section, then a unilateral stress, Z, acting on two opposite sides of the cube and in, say, the z-direction, will be carried equally by the M chains. Hence, it is sufficient to consider the behavior of one chain under the action of a stress Z/M, and derive an equation of state for this simple chain.

Our problem is closely related to the derivation of an equation of state for a gas, and may be solved in an entirely similar fashion.

A chain of rubber molecules may be considered as analagous to a gas whose molecules are tied together by flexible strings. The strings must be flexible, because of the free rotation. If there were no strings, the heat motion would cause an expansion of the gas. The presence of the string will cause a coiling-up of the resulting chain in a random fashion. To mention a simple illustration, an actual piece of flexible string thrown into the air will come down in a coiled form rather than in a straight form. *This is true because of the greater probability of coiled configurations, in contrast to the least probable straight configuration.*

5. Explanation of the Anomalous Thermoelastic Behavior

The simple bundle model explains qualitatively the anomalous thermoelastic behavior of rubber. It explains, for example, the fact that heat is developed when rubber is stretched rapidly. In the unstretched sample, the rubber chains are coiled up, corresponding to a most probable state. Work is necessary to bring the rubber chains from their more probable coiled-up form into a less probable straightened-out form. This work in turn is transformed into heat. The rise of temperature observed on fast stretching of rubber is entirely similar to the temperature rise produced when a gas is suddenly compressed.

6. Results of the Statistical Treatment of the Bundle Model

The equation of state $Z = Z(L, T)$, where Z = force per unit original cross-section, for the bundle model, may be derived by a statistical method entirely similar to that used to derive the equation of state $p = p(V, T)$, of a gas. Neglecting the contribution to the stress due to the internal energy, we obtain:

$$Z = Z(S) = KTL \tag{3}$$

Here K is a constant, and L may be interpreted as the relative length: *i.e.*, the extended length divided by the original length.

This equation explains *qualitatively* the three thermoelastic effects exhibited by rubber and the stress-strain curve. However, it holds only for medium extensions.

EQUATION 3 yields the observed rise in temperature on fast stretching.

EQUATION 3 gives for the linear thermal expansion coefficient α (parallel to the direction of the stress),

$$\alpha_{\parallel} = -1/T, \tag{4}$$

i.e., the linear thermal expansion coefficient of *stretched* rubber has exactly the same magnitude (in the approximation of EQUATION 3) as the cubic expansion coefficient of an ideal gas, but it has a negative sign. Thus, EQUATION 3 explains why rubber stretched isotonically contracts visibly when heated. $\alpha_{\parallel}$ (in the approximation of EQUATION 3) is quite unrelated to the linear thermal expansion coefficient of *unstretched* rubber, which is of the order of 2.10^{-4}.

For rubber stretched isometrically, EQUATION 3 gives the (proportional) increase of stress with temperature.

The upward curvature of the stress-strain curve, as shown in FIGURE 4, may also be explained on the basis of the bundle model. For larger extensions, EQUATION 3 has to be replaced by another one containing the so-called Langevin function (cf. FIGURE 20), which occurs in Langevin's theory of paramagnetism and Debye's theory of polar molecules. The orientation of the chains by the stress is very similar indeed to the orientation of dipoles by fields.

Thus, the bundle model does explain the anomalous thermoelastic properties of rubber. It also explains the shape of the stress-strain curve at medium and larger extensions. The bundle model has, however, the following shortcomings: (a) It is essentially a one-dimensional model, stretchable only in the direction of the axis of the bundle; also, $\alpha_{\perp}$, the linear thermal expansion coefficient perpendicular

to the direction of the stress, cannot be computed. (b) It does not hold for small extensions and for compression. Experimentally, it is found that Hooke's law holds in the limiting case of very small extensions. The bundle model gives $Z > 0$ for $L = 1$, and $Z = 0$ for $L = 0$: *i.e.*, zero volume for the unstretched state. (c) It does not explain the thermoelastic inversion point.

A somewhat more detailed discussion of the material of this section is given in SECTION 1 of the Appendix. In particular, EQUATION 3 is derived there by statistical methods.

7. The Network Model

We know that, even in raw rubber, we do not have a bundle of isolated long chain molecules. Rather, the long chain molecules must form a coherent mass by some sort of cross-linking, for there must exist strong forces between the molecules to prevent their slipping past each other. Otherwise, raw rubber could not be elastic, but would show as much plasticity as beeswax and chewing gum, substances which show no appreciable retraction after the release of a stretching force.

In the vulcanization process, the cross-links already present in raw rubber are reinforced and new cross-links created. Cross-links may be primary and secondary bonds. This general picture of the network structure of rubber is sufficient for the theory developed here.

We wish to emphasize that the existence of the network is not derivable (at least not at present) from sub-microscopic or other more direct evidence. The existence of a network must be postulated to account for the physical and chemical changes accompanying vulcanization.

For soft rubber, even after vulcanization, there are only a few cross-links per chain. Although a translation of the chains with respect to each other is suppressed, there is enough freedom of motion of the chains to make possible a free rotation around single bonds in the sense defined above.

We take as our model a region crossed by a complicated network of chains, very irregular in detail, but homogeneous on the average. The chains may run between perpendicular plane boundaries of the volume of the rubber over which their ends will be uniformly distributed. The flexible chain molecules are bonded into the network by steric forces.

One action of the network will be a pull, the development of a contracting force due to the tendency of the chains of the network to coil up because of free rotation. This contracting force would lead

to a collapse of the network, if there were not present another action of the network exactly balancing the contracting force.

The second action of the network is due to the effects of molecular bombardments on the bounding surfaces. The chains jerk at the boundaries of the bulk material. In addition, there may be inert clusters of chains present (a sol component) which are not connected with the network. Both the molecular bombardments on the bounding surfaces and the sol component will lead to the production of a pressure as in an ordinary liquid.

In our model, we separate rather sharply the two actions of the network: the contracting pull of the network due to the tendency of the chains to coil up, and the hydrostatic pressure, due to the sidewise pushes of the chains. We compute the first action, the pull of such a network, directly, but represent its second action, the production of internal pressure, simply by filling the model with a fictitious* liquid, the "rubber liquid," which, to a good approximation, one may take as incompressible. Any surface of the model must, of course, be in equilibrium under all the forces which act on it: the pull of the network, the push of the hydrostatic pressure, and any external forces.

8. Results of the Statistical Treatment of the Network Model

The bundle model neglects the volume-filling properties of rubber. It neglects, in the terminology of SECTION 7, the internal pressure. The external force (stress), Z, is then equal to the inward pull due to the contracting force which is, according to EQUATION 3, equal to KTL. According to the network model, however, the external force is equal to the *difference* between the inward pull due to the contracting force *and* the outward push due to the internal pressure. This push is equal to $KT(1/L^2)$. The stress-strain relation for the network model is, therefore:

$$Z = KT(L - 1/L^2). \tag{5}$$

The constant, K, is determined by the molecular weight, M, of the chains between two junctions of the network. A reliable establishment of the correlation between K and M is possible only on the basis of a theory of network formation. Such a theory was recently put forward by James and the author. The relation $K \backsimeq \frac{1}{2}\frac{\rho}{M} \cdot R$, holds to the order of magnitude (ρ:density, R gas constant); K is thus pro-

* This "liquid" is fictitious as far as it simulates effects of steric forces and molecular bombardments, but more nearly real as far as it represents a sol component.

portional to the number of cross-linkages. For comparison of EQUATION 5 with the experiment, compare FIGURE 21.

The three-dimensional character of the network model permits the computation of the *linear thermal expansion coefficient perpendicular to the direction of the stress.*

$$\alpha_{\perp} = +\ 1/2T, \tag{6}$$

in the approximation of EQUATION 4. *Thus, stretched rubber shows a marked thermal anisotropy.*

For larger extensions, one obtains, instead of EQUATION 5,

$$Z = K'T[\mathfrak{L}^{-1}(L \cdot \kappa) - 3\kappa\ L^2], \tag{7}$$

where $\mathfrak{L}\ (x) = \coth\ x - 1/x$ is the Langevin function, mentioned previously, and κ and K' are constants. For small extensions, EQUATION 7 reduces to EQUATION 5.

The *thermoelastic inversion* can be explained by taking account of the cubic thermal expansion of unstretched rubber, characterized by a coefficient of cubic expansion, α, defined by the equation:

$$V(T) = V(T_o)\ [1 + \alpha(T - T_o)]. \tag{8}$$

Qualitatively, the thermoelastic inversion is simply an expression of the competition of the outward push due to thermal expansion and the resulting inward pull due to the difference between contracting force and effect of internal pressure. It is clear, then, that there must be a critical extension, the thermoelastic inversion point, where the two competing forces cancel each other. *Quantitatively,* the effect of thermal expansion may be introduced as an additional internal pressure. We obtain:

$$Z = KT\left[L - \frac{1 + \alpha(T - T_o)}{L^2}\right], \tag{9}$$

which gives $Z = 0$ for $L = 1 + \alpha/3\ (T - T_0)$, as it should, while EQUATION 5 gives $Z = 0$ for $L = 1$.

A more detailed discussion of the consequences of EQUATION 9 is deferred to SECTION 3 of the Appendix.

In EQUATION 9, a part of the stress is due to internal energy connected with thermal expansion. This part plays an important role at small extensions only (in the neighborhood of the thermoelastic inversion point), but can be neglected at larger extensions.

Thermal expansion may similarly be introduced into EQUATION 7, since it affects the internal pressure only; this pressure has the same value in EQUATIONS 7 and 5. From the resulting equation, one can

compute $Z(U)$ and $Z(S)$ (cf., for the details, SECTION 6 of the Appendix) and compare them separately with experimentally obtained $Z(U)$ and $Z(S)$. FIGURE 7 shows this comparison. The experimental data were obtained by Anthony, Caston, and Guth. The overall agreement between theory and experiment is fairly good, except at the lowest extensions where van der Waals forces, neglected in the theory, may play a role.

It should be pointed out that the theory of James and Guth provides

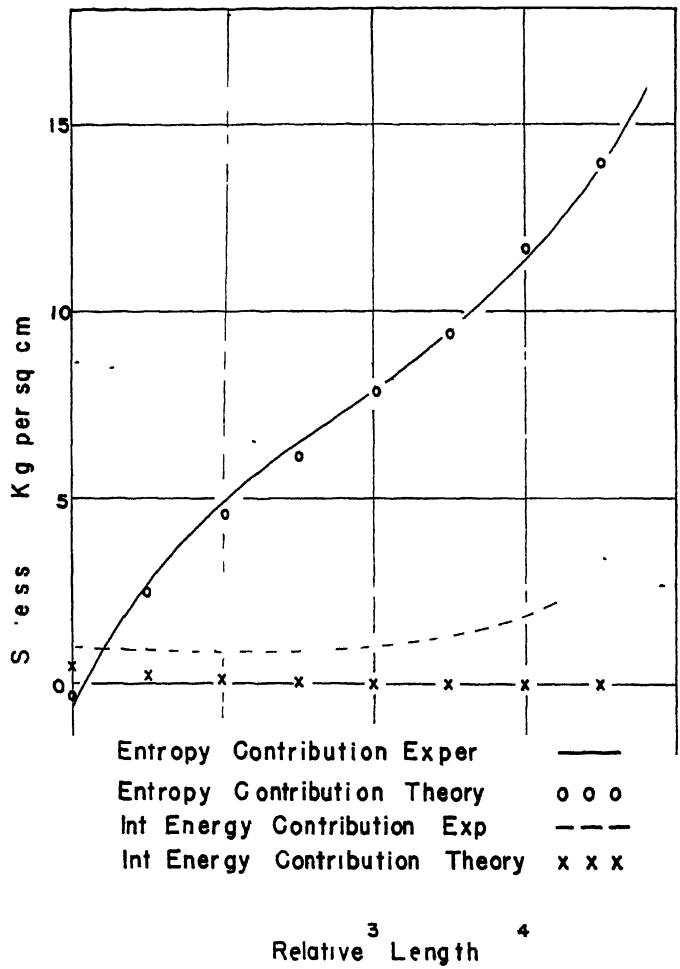


FIGURE 7. Internal energy and entropy contributions to the stress for an unaccelerated pure gum compound. The circles are theoretical values according to EQUATION 46a.

a general basis for the explanation of all physical properties of stretched rubber.

The theory developed by James and Guth does not consider crystallization phenomena, which are not essential for the occurrence of rubberlike elasticity. In fact, it is known that Buna S (GR-S) type synthetic rubber does not crystallize and still exhibits the characteristic S-shape in its stress-strain curve. The unaccelerated pure gum stock of FIGURE 7 also shows no crystallization up to about 400 % extension, and still yields an S-shaped stress-strain curve over this region.

A more detailed treatment of the material of this section is contained in SECTIONS 2, 3, 4, 5, and 6 of the Appendix.

9. General Occurrence of Rubberlike Materials

According to the principles set forward in our theory, all long chain compounds with free rotation will exhibit rubberlike elasticity above their respective melting points. Conversely, any compound exhibiting rubberlike elasticity will be a long chain compound possessing free rotation. The long chain compounds have a melting point or melting range above which there is a free rotation. Whether this melting point is below or above room temperature depends on the competition between free rotation and molecular forces. The free rotation may be started either by raising the temperature or by introducing a solvent through the process of swelling or plasticizing. A survey of rubberlike materials was given by the author in another paper.* Here we restrict ourselves to a few examples:

(a) Guttapercha, the trans-isomer of rubber, shows rubber-elasticity above 70° C.

(b) The inorganic substances, sulfur, selenium, phosphorus chloronitride, and sulfur trioxide, possess elastic modifications at higher temperatures and under special circumstances.

(c) Synthetic elastomers, like neoprene, Butyl rubber, thiokol, and the butadiene-type of substances;

(d) polystrene, polyindene and all allied substances;

(e) cellulose derivatives;

(f) nylon;

(g) biological materials, like gelatine and glue, wool and silk, hair, chromosomes, and the *muscles*.

* Publication No. 21 of the American Association for the Advancement of Science. Cf. reference 1.

10. Summary of Rubberlike Elasticity in the Steady State

Fundamental aspects of rubber-elasticity are the thermoelastic effects and the S-shape of the stress-strain-compression curve of rubber. There are three thermoelastic phenomena for elastomers:

1. They exhibit a rise in temperature on fast stretching.
2. If stretched by a constant load (*i.e.*, isotonically), they contract visibly when heated.
3. The stress in an elastomer kept at constant extension (*i.e.*, isometric stretch) increases with rising temperature.

These three effects are correlated by thermodynamics.

The theory of James and the author explains quantitatively all these fundamental properties of rubber-elasticity. It explains observed finer details of thermoelasticity. The change in temperature on fast stretching is negative (cooling) for low extensions, passes through a minimum, becomes zero (adiabatic thermoelastic inversion point) and finally takes on positive values (heating) for increasing extensions. It explains the observed thermal anisotropy of isotonically stretched rubber. The linear thermal expansion coefficient is negative in the direction of the stress, but positive in the direction perpendicular to the stress. Both for isotonic and isometric stretch, there exists a thermoelastic inversion point which is half of the value of the critical extension for the adiabatic thermoelastic inversion point. The visible contraction of isotonically stretched rubber, when heated, is due to the fact that the negative linear thermal expansion coefficient of stretched rubber has about the same magnitude as the (cubic) thermal expansion coefficient of a gas. The thermoelastic inversion points are determined in a simple manner by the cubic thermal expansion coefficient of unstretched rubber. The S-shape of the stress-strain (compression) curve is also explained by the theory.

The thermoelastic effects, with the exception of the inversion phenomena, may be explained on the basis of the bundle model. In this model, the bulk rubber is assumed to consist of a bundle of parallel strings built up from flexible (free rotation) long chain molecules. However, only the new network model of James and the author can explain the stress-strain curve at low extensions and its continuation for compression. This model is also needed to explain the thermal anisotropy of stretched rubber. Finally, the inversion phenomena may be explained by including in the theory the cubic thermal expansion of unstretched rubber.

11. Crystallization

We have referred already to crystallization, in connection with the thermoelastic and elastic behavior of rubber at high extensions. It remains to put forward simple pictorial representations for the crystallization process.

Above a certain transition temperature T_c (more accurately: range of transition temperatures), unstretched rubber forms an amorphous network, as pictured in FIGURE 8. Below T_c, it is partially crystallized,

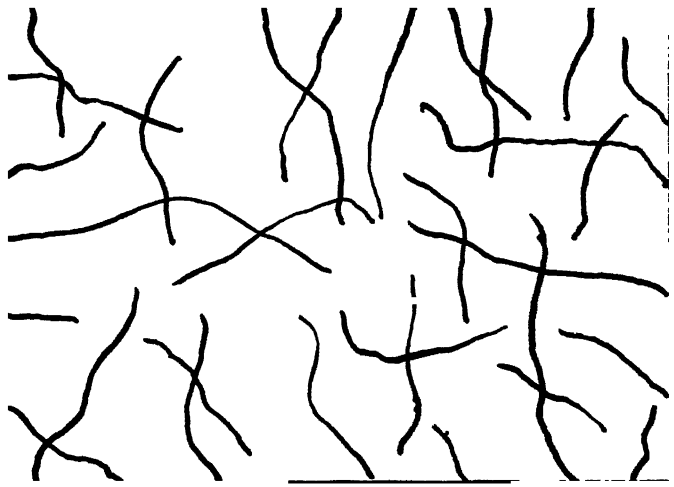


FIGURE 8. Random network of rubber chain molecules.

i.e., there is a parallelization of the chains in crystallites or micelles. These crystallites are randomly oriented, as shown in FIGURE 9. They are estimated to be 3-600 Å long, while the chains themselves may be as long as 10,000 Å or more. Therefore, an average chain must pass through many crystallites. Thus amorphous regions connect the crystallites. The extent of crystallization depends upon the external variables (stress and temperature) and type of rubber. Amorphous rubber may be crystallized by stretching it sufficiently without lowering the temperature. The crystallites, in this case, are lined up in the direction of the stress, as shown in FIGURE 10.

As previously mentioned, the upward turn of rubber stress-strain curves need not necessarily be due to crystallization. Orientation of the chains also leads to an S-shape, as proved by the statistical theory of James and the author. However, crystallization enhances the up-

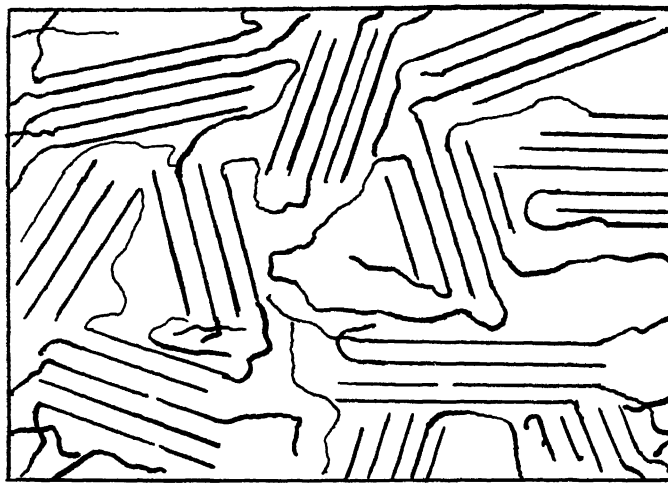


FIGURE 9. Crystallization of unstretched rubber. The regions with parallel chains are the crystallites or micells.

ward turn. The stiffening effect of crystallization is similar to the stiffening effect of anisotropic fillers in rubber. The forces parallelizing the chains in the crystallites are stronger than the forces between chains in the amorphous region. Consequently, the crystallites may be considered as fillers whose rigidity exceeds the rigidity of the sur-

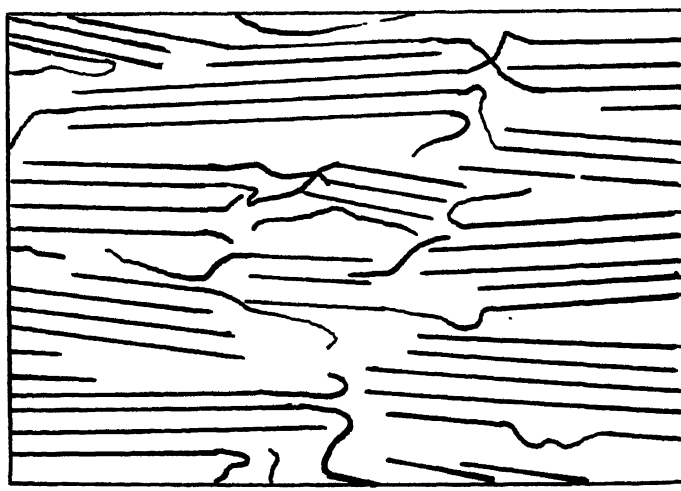


FIGURE 10. Crystallization of stretched rubber. The micells are oriented in the direction of the stress.

rounding (amorphous) rubber matrix. Then, at least qualitatively, the theory of the stiffening effect of anisotropic fillers, developed by the author,* applies. Of course, the number of crystallites, *i.e.*, their concentration, changes with stretch, while it is a constant for the case of actual fillers.

It is not possible to develop a theory of the crystallization of rubberlike materials as general as our theory of rubber elasticity. The magnitude and geometrical distribution of the intermolecular forces regulating crystallization will depend strongly upon the chemical composition of the elastromer (cf. ref. 3, pp. 279–285).

12. Vulcanization of Stretched Rubber

Rubber vulcanized in the stretched state gives, on further stretching, isometrics with a negative slope. Cross-linking of the oriented chains apparently prohibits free rotation to a large degree. Rubber shows a normal thermoelastic behavior under these conditions. Vulcanization introduces, of course, a stiffening of the (previously unvulcanized) material.

B. TIME EFFECTS FOR RUBBER[7]

1. Stress Relaxation and Creep

Stress relaxation is the decay of stress with time, if a material is stretched to a constant extension. It occurs for every material. The peculiarity of rubber shows up, however, in the quantitative stress-time relationships. A typical stress-relaxation curve for rubber is shown in FIGURE 11. Experiment shows that the stress, as a function of extension ϵ, temperature T, and time t, may be factored as follows:

$$Z(\epsilon, T, t) = F(\epsilon, T) \cdot G(t, T). \tag{10}$$

This factorization implies that stress-time curves for various extensions are similar; *i.e.*, they differ by constant factors only. Stress-strain curves for various times obtained from stress-time curves by cross-plotting are also similar.

The first factor in EQUATION 10 is the stress-strain curve (EQUATION 5) of the steady-state theory, while the second factor represents the rate process of stress decay. For certain cases in which more and more cross-linkages are ruptured with time, the factor G may be interpreted as the time dependent number of cross-linkages.

* Guth, E. J. Appl. Phys. 16: 20. 1945.
[7] cf. Guth, E., P. E. Wack, & R. L. Anthony. J. Appl. Phys. 17: 347. 1946.

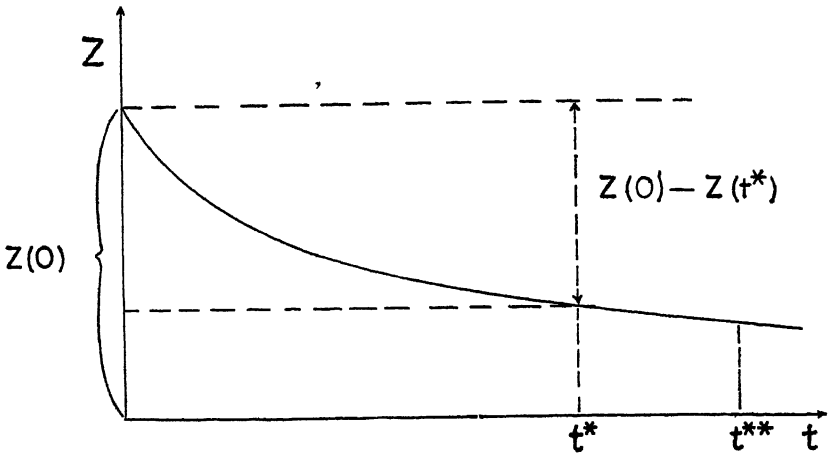


FIGURE 11. Stress relaxation. t* indicates the time after which isometrics are taken in the relaxation method of obtaining isometrics.

Creep is the increase of strain with time, if a material is deformed by a constant stress. Creep is shown schematically in FIGURE 12 for *short times,* while the extension, as a function of stress, temperature, and time, may be obtained from EQUATION 10 by solving for ε. It is seen then that creep curves taken with various values of stress are similar (differing only by a multiplying factor) only if the strain is very small so that Hooke's Law holds. For longer times, stress-relaxation and creep are connected by a formalism which gives time depend-

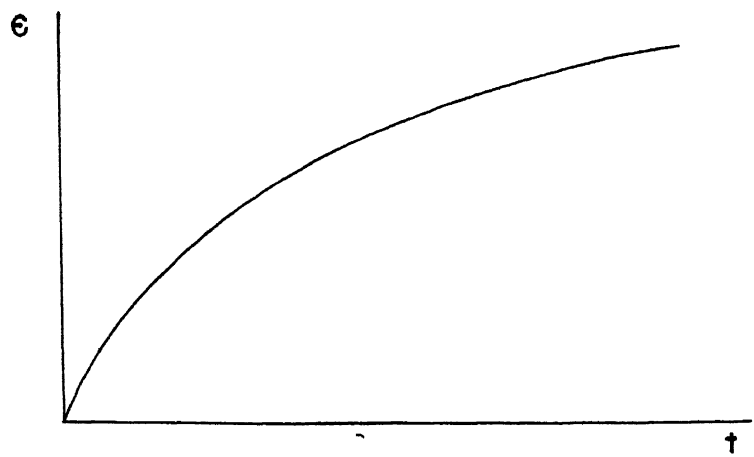


FIGURE 12. Creep.

encies in a manner similar to that whereby thermodynamics gives temperature dependencies of various effects. The influence of speed on stress-strain curves obtained at a constant rate of extension or stress may be obtained from the same formalism.

2. Relaxation Method of Obtaining Isometrics

Because of stress relaxation, stress-strain curves show the expected dependence upon temperature only if they are taken so fast that no appreciable change in the factor G in EQUATION 10 occurs. In practice, it is difficult to obtain the stress-strain-temperature data with the required speed. In a more convenient method, the sample is relaxed at a fixed temperature, T^*, until a time, t^*, is reached (cf. FIGURE 11) where the stress decays very slowly with time. Then the temperature is lowered and the corresponding change in stress observed. The time t^* is chosen so that, during the single interval $(t^{**} - t^*)$ needed for lowering the temperature in the experiment, the change of stress with time is within the experimental error. Thus, isometrics are obtained directly. The duration, t^*, and the temperature, T^*, of the pre-relaxation can change the isometrics by numerical factors only, because of the experimentally verified relations:

$$Z(\epsilon, T, T^*, t^*) = F(\epsilon, T), G(T^*, t^*). \tag{11}$$

The validity of this factorization implies the similarity of certain curves, as in the previously discussed case of relaxation. This relation should be checked for each material to which the method is applied.

3. Retraction of Stretched Rubber

The retraction of stretched rubber has been studied recently[8] at Notre Dame, both by the kymograph and by high speed photography. Kymograph records (elongation-time curves) for natural and for Butyl rubber are shown in FIGURES 13 and 14. It is seen that, for natural rubber, the speed of retraction (slope of the elongation-time curve) is a constant. The curvature of the elongation-time curve at the beginning of the retraction is due to the weight of the stylus and of the rubber used for clamping before the stress is released. For Butyl rubber, on the other hand, the velocity goes from a maximum to zero.

[8] **Mrowca, B. A., S. L. Dart, & E. Guth.** J. Appl. Phys. **16**: 8. 1945.

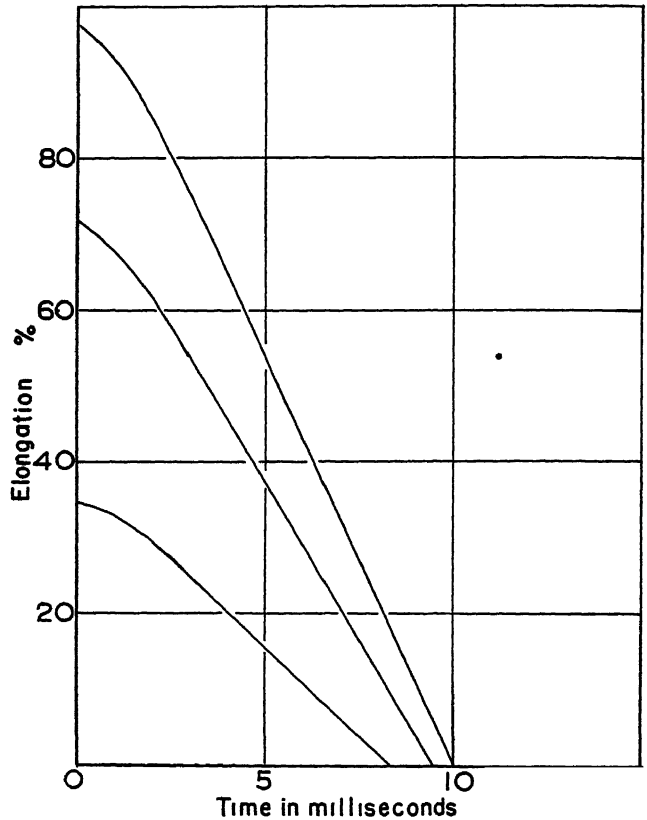


FIGURE 13 Elongation-time curves for the free retraction of stretched natural rubber

Theory[9] shows (cf. SECTION 8 of Appendix) that the (constant) speed of retraction for natural rubber is determined by the inertia of the elements of the sample. This is similar to the inertia effect observed when a train starts to move; the last car moves only after a "wave" passes down to it along the couplings of the cars. The elastic wave pulse passing down the rubber sample may be seen very clearly in high speed photographs. For Butyl rubber, a combination of inertia and internal friction has to be assumed, the latter being responsible for the decrease of the velocity to zero (cf. EQUATION 62a). The quantitative treatment of this process is indicated briefly in SECTION 8 of the Appendix.

[9] James, H. M., & E. Guth. Phys. Rev. 66: 33. 1944.

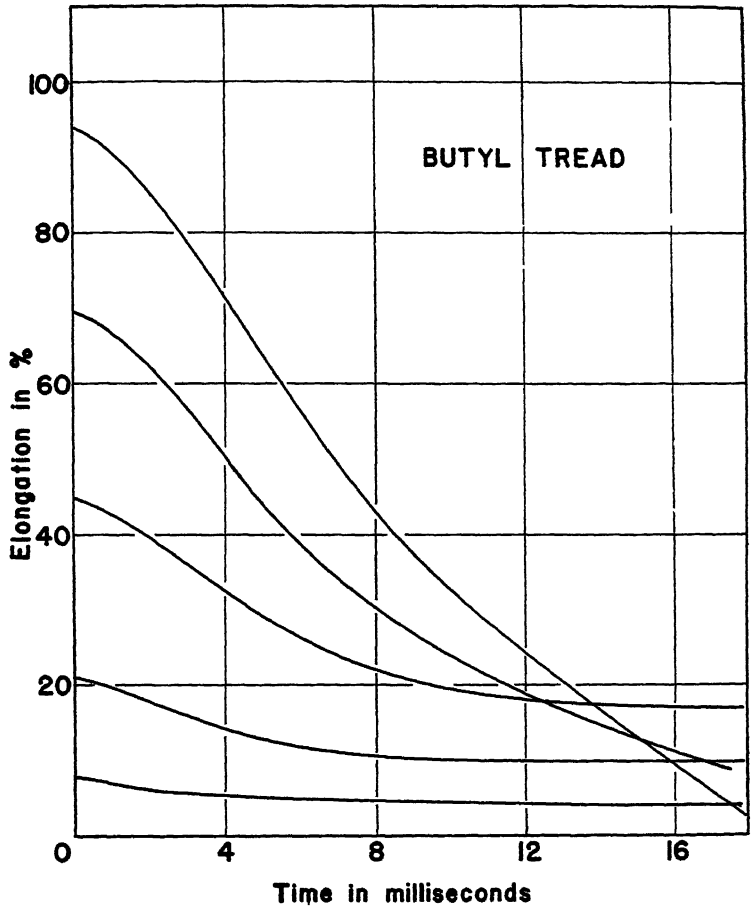


FIGURE 14 Elongation-time curves for the free retraction of stretched Butyl rubber.

C. ELASTIC PROPERTIES OF RESTING MUSCLE

1. Stress-Strain Curve

The stress-strain curves of resting muscle and of rubber are significantly different (cf. FIGURES 15a and 15b). Examination of FIGURES 15a and 15b shows that, for extensions exceeding ε^* (FIGURE 15b), the rubber stress-strain curve has a characteristic similar to that of resting muscle (FIGURE 15b). In other words, muscle corresponds to rubber stretched out so much that the chain molecules are markedly oriented. As we have already noted (cf. SECTIONS A 6 and 8), the up-

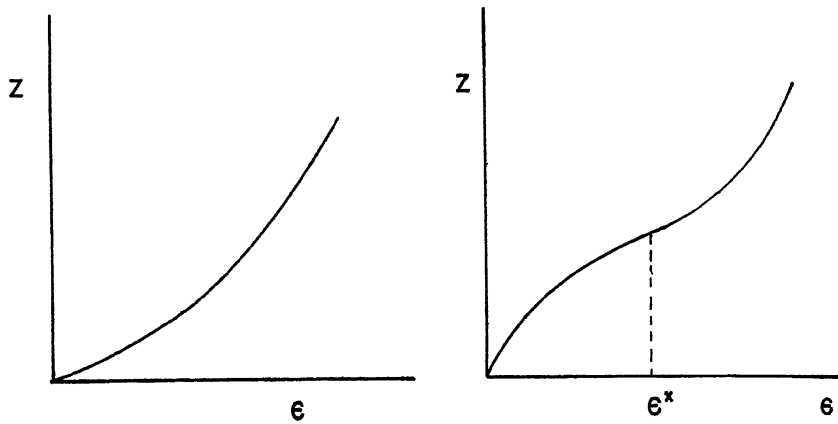


FIGURE 15a. Stress-strain curve for resting muscle (schematic).
FIGURE 15b. Stress-strain curve for rubber (schematic).

ward turn of the rubber stress-strain curve (above ε^*) may be explained by a marked orientation of the molecular chains. Crystallization, if it occurs, greatly enhances the upward turn, but is not necessary for its explanation. Fenn's statement (*loc. cit.*, p. 469) that, "as the stretch increases, new structures within the muscle are called upon to offer resistance," may be made more precise by substituting, for "new structures," crystallized or oriented regions. X-ray and birefringence data (cf. the paper by E. Fischer* and the results of Treloar's (Trans. Faraday Soc **37**: 84. 1941.) work on birefringence of crude natural rubber, showing an interesting similarity with the corresponding behavior of resting muscle) would seem to indicate that crystallization plays an important role for muscle. This is not necessarily true for rubber.

Howell's (otherwise quite good) account of muscle physiology in his textbook[10] is somewhat misleading, concerning the comparison between muscle and rubber (cf. his Figure 4). Rubber does *not* obey Hooke's Law, except for very small (1–2 %) extensions. At 5 % extension, there is a 5 % deviation from Hooke's Law, and so on. In Howell's Figure 5, a stress-strain curve for muscle, due to Marey, is presented. Up to the point *a*, this curve shows the behavior schematically represented in our FIGURE 15a. Whether the part beyond *a* has significance (and is due to plastic flow, as Howell believes), can be stated only after further investigation, for the muscle may start to fall apart[11] beyond point *a*.

* P. 783 ff.
[10] **Howell, W. H.** A Text-Book of Physiology. 13 ed. Saunders, Philadelphia. 1937.
[11] This was pointed out to the author by **A. S. Gilson.**

The work of Banus and Zetlin, and particularly that of Ramsey and Street, indicates that the tension of resting muscle is due entirely to the connective tissue. Histological work has shown that the sarcolemma is analogous to a knitted stocking composed of a network of fibers. Whether muscle contains a molecular network similar to that of rubber, has not been definitely proved. However, the presence of such a network is made very probable by the existence of a long range of reversible elasticity for muscle.

2. Thermoelastic Effects

Heidenhain, in 1863 (*i.e.*, only four years after Joule's publication on rubber), showed that resting muscle generates heat when it is stretched rapidly, and cools when it is released. Recently, Heidenhain's discovery was confirmed by Hill and Hartree, and by Feng. Experimentally, these men found that (for muscle) ΔT first increases as the extension is increased, then passes through a maximum and, for still greater extensions, becomes negative. This behavior is shown in FIGURE 16. Using essentially our EQUATIONS 9a and 12a, Hill and Hartree substituted the empirical values of ΔT, and thereby computed the linear coefficient of thermal expansion for muscle. The occurrence of an inversion point in FIGURE 16 seems to indicate the

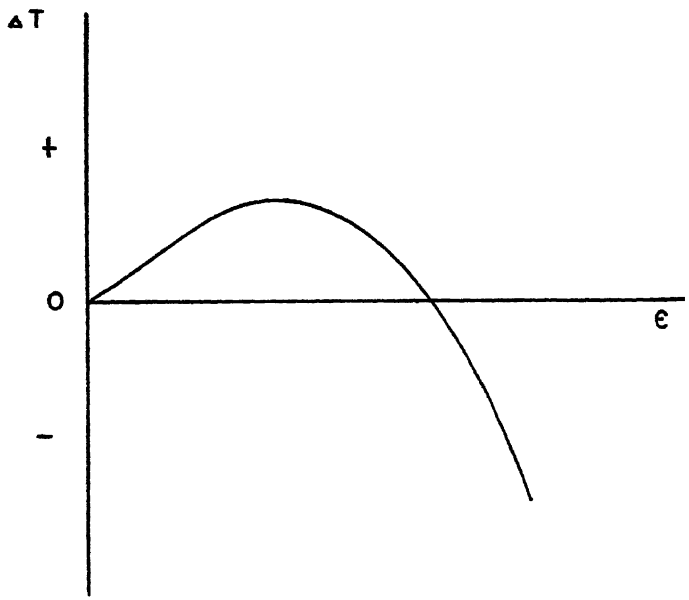


FIGURE 16. Change in temperature on fast stretching of muscle (schematic).

completion of crystallization (and of molecular orientation) on stretching. Beyond that point, muscle exhibits a normal thermoelastic behavior. The results obtained by Dart, Anthony, and Guth, and Dart and Guth, for the rise of temperature produced by the rapid stretching of natural and synthetic rubbers, also show the presence of this inversion point.

Isotonic contraction, obtained by heating stretched resting muscle, has been studied by Wöhlisch and his associates. In agreement with Hill and Hartree, and Feng, they found that $\alpha_{\parallel}$ (cf. EQUATIONS 29a and 40a), the linear coefficient of thermal expansion parallel to the direction of the stress, is negative first, passes through a minimum, and then increases, passing through zero at the thermoelastic inversion point, and becomes positive thereafter (cf. FIGURE 17).

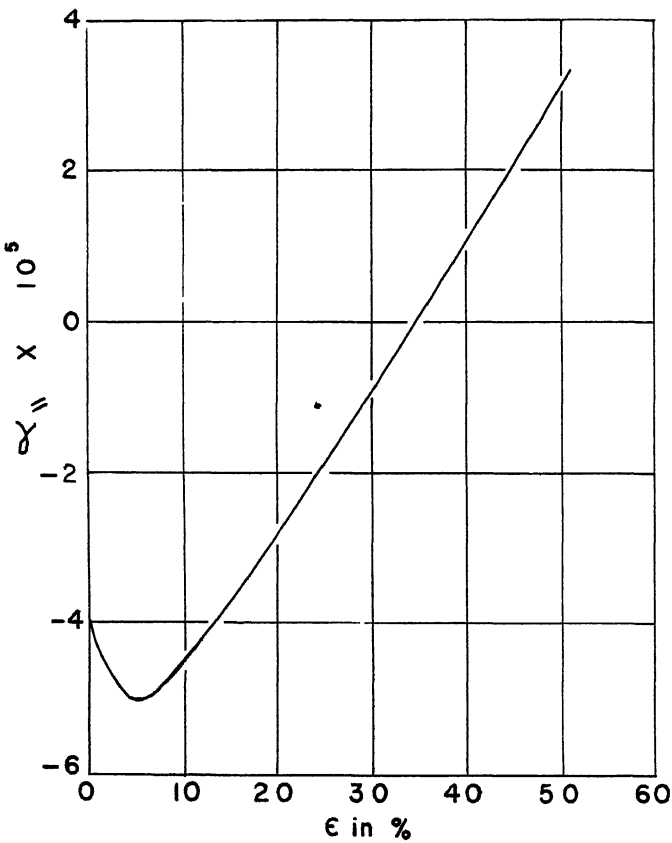


FIGURE 17. $\alpha_{\parallel}$ for resting muscle (schematic).

Isometrics have been investigated by Meyer and Picken, and lately by Wöhlisch *et al.* FIGURE 18 shows the result of the latter for *ligamentum nuchae.* One sees clearly that muscle corresponds to rubber partially crystallized as the result of stretching. Wöhlisch *et al.* also checked the consistency of their data by direct measurements of the rise of temperature, employing, in effect, our EQUATION 9a of the Appendix. For a "good" rubber, $Z(U)$ is usually small compared with $Z(S)$. For muscle, the two are of the same order of magnitude, showing clearly the significance of molecular interactions besides the pure entropy effect. A similar result was obtained by Wöhlisch *et al.* for skeletal muscle. (None of the authors has carried out quantitative

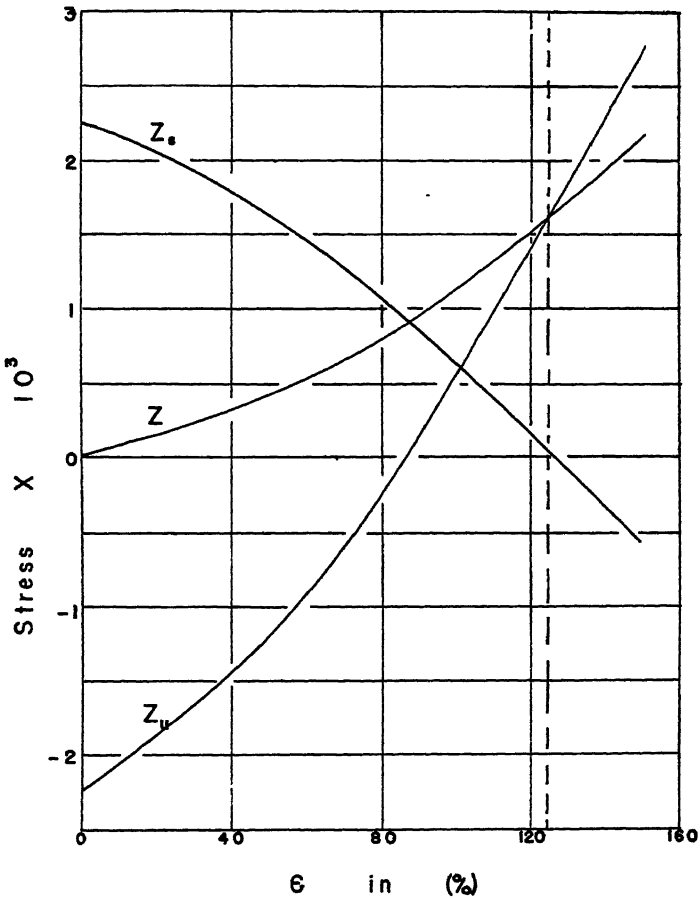


FIGURE 18 Resolution of the stress into its components, $Z(S)$ and $Z(U)$, due to internal energy and entropy, respectively for *ligamentum nuchae.* This substance is simpler and more similar to rubber than skeletal muscle.

work on both stress-strain curves and at least two thermoelastic effects. None of them investigated the time effects which enter into the determination of isometrics; for instance, the work by Meyer and Picken is not sufficiently complete.)

Summarizing, the elastic and thermoelastic behavior of muscle corresponds closely to that of rubber partially crystallized through stretch. The concave (to the ordinate axis) character of the stress-strain curve, the rise in temperature on fast stretching, the isotonic contraction on heating, and the resolution of stress into its additive components due to internal energy and entropy, respectively, all point in the same direction.

3. Stress-Relaxation of Resting Muscle

The stress-time behavior of active muscle was recently studied in detail by Gilson, Walker, and Schoepfle,[12] and by Schoepfle and Gilson.[13] The stress-relaxation of resting muscle, however, does not seem to have been studied. At the suggestion of the author, Drs. Gilson and Schoepfle kindly performed a stress-relaxation experiment on the sartorius muscle of a bullfrog. The initial length of the sample was $l_0 = 61$ mm., and it was stretched to 76.5 mm. The decay of the stress is shown in FIGURE 19 in semilog representation. It is seen that, after an initial fast decay, the curve is a straight line: *i.e.*, the stress decays thereafter as an exponential function of the time. The initial fast decay is probably connected with the temporary breaking of van der Waals' bonds: *i.e.*, it is a primary relaxation. The exponential part is probably connected with the breaking of chemical bonds: *i.e.*, it is a secondary relaxation. After three hours, the strain was released and the sample's length took on rapidly the value 71.5 mm. and stayed there. The rapid recovery corresponds, of course, to the fast initial decay, while the permanent set corresponds to the exponential decay.

A further study of stress relaxation, in conjunction with studies of the creep and of the influence of speed upon stress-strain curves, would be very useful. In addition, the retraction of resting muscle could give interesting results. The relaxation method of obtaining isometrics would give useful clues as to the nature of the elasticity of resting muscle.

[12] **Gilson, A. S. Jr., S. M. Walker, & G. M. Schoepfle.** J. Cell. & Comp. Physiol. **24**: 185. 1944.

[13] **Schoepfle, G. M., & A. S. Gilson, Jr.** Ibid. **26**: 119. 1945.

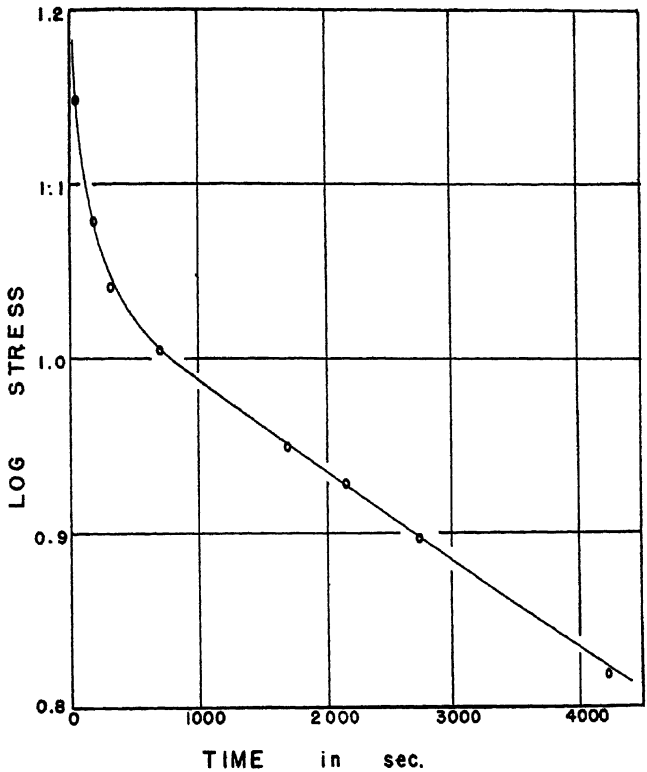


FIGURE 19. Stress-relaxation of frog sartorius muscle.

D. ELASTIC PROPERTIES OF ACTIVE MUSCLE

1. Structural Aspects

The elastic and thermoelastic behavior of the resting muscle shows great similarity to the corresponding properties of rubber, as we have seen. However, even for resting muscle, the similarity with rubberlike materials represents one aspect of the problem only. The release of metabolic energy and the change in pH on stretching of the resting muscle show clearly the limitations of that similarity. The "non-rubberlike" properties of muscle due to the ionic environment and accompanying chemical changes may outweigh the rubberlike properties in importance. For the active muscle, ionic environment and chemical changes are much more important aspects than for the resting muscle.

Nevertheless, comparison of the active muscle with rubber can help a little in the clarification of the fundamental problems of the *elasticity* of the *active* muscle. The structure of muscle, as explored by electron microscope and x-ray studies, is surveyed in the excellent paper* by Schmitt, Bear, Hall, and Jakus. For resting muscle, it is possible to ascribe the observed elasticity and thermoelastic behavior entirely to the connective tissue, without knowing the structure of the myofibrils and how they are bound to the connective tissue. For the active muscle, however, these two questions are of paramount importance. A myofibril consists of fine filaments, 50–250 Å in width, which extend continuously through *A* and *I* bands. The main difference between *A* and *I* bands is that the *A* bands contain relatively dense material. The reversal of striation on contraction (Jordan) was checked; this reversal means the movement of the *A*-substance from the *A* bands toward the *I* bands. The continuous extension of the filaments through *A* and *I* bands, discovered by Schmitt *et al.*, and the reversal make it possible to consider *A* and *I* bands as essentially of the same structure. On contraction, the filaments themselves shorten rather than coil up. The myosin filaments are the smallest contractile units thus far observed. *No direct information is available, at present, on the microstructure of the myosin filaments.* We can only make some guesses. The simplest assumption seems to be that the myosin filaments form a sub-microscopic stationer's rubber band which is *stabilized* in a stretched state. With this picture, no essential difference between myosin filaments and connective tissue has to be assumed as far as their network structure is concerned, *i.e.*, the folding of chains may take place inside the filaments. In order to explain the anisotropy of the *A* and the isotropy of the *I* bands, we may assume that, in the resting state, the former are stabilized at a more extended state than the latter. This does not exclude the possibility that the *I* bands may be actually somewhat compressed in the resting state.

How are the myosin filaments stabilized in a stretched state? The recent investigations of the Szent-Györgyi[14] school (cf., also, the paper by Steinbach†) show that the fibrous protein, previously called just myosin, actually consists of two components, actin and myosin. Actin is rigid and may prop up the flexible myosin, if more or less strong cross-linkages bind together these two types of proteins. On

[14] **Szent-Györgyi, A.** J. Colloid Sci. 1: 1. 1946.

* Pp. 799–812.

† P. 849ff.

stimulation, these cross-linkages are dissolved, or the actin may become flexible. All this, of course, is merely conjecture.

2. Coupling Between Sarcolemma and Molecular Matrix

The next problem, the *mechanical coupling* between resting (*i.e.*, sarcolemma) and active muscle, is rather unclear at present. This is particularly unfortunate, since it makes uncertain the interpretation of any observed elastic or thermoelastic property of the active muscle. The possibility of extraction of myosin filaments by dilute KCl solution (Greenstein and Edsall) shows that, in the resting state, these filaments are not linked too strongly to the muscle matrix. This method of extraction leads to a wide distribution of lengths, as indicated by Edsall and extended by Schmitt *et al.* (*loc. cit.*). The fact, established most directly by the "pinching" experiment of Ramsey and Street, that the myofibrils do not contribute measurably to the tension of stretched resting muscle, points in the same direction. Hence, no conclusions may be drawn about the myosin filaments from the elastic behavior of resting muscle.

This weak (if any) binding of the myosin filaments to the sarcolemma in the resting muscle does not necessarily involve an equally weak binding in the active muscle. The stimulus may cause the formation of strong bonds between myosin filaments and sarcolemma. In another connection, Astbury suggested that, on stimulation, a sort of vulcanization of the muscle may take place. It would be of great interest, in this connection, to repeat the Ramsey-Street experiment by pinching the muscle before it is stimulated.

Two interesting suggestions were made recently to explain the transmission of force from the myosin filaments to the sarcolemma, without assuming end-to-end bonds between them. Bull[15] suggested a sort of "fluid drive" coupling for the transmission of longitudinal tension from the myosin filaments to the sarcolemma. It was pointed out, however, both by the author and Dr. J. T. Edsall, that such a mechanism is somewhat hard to understand. Schoepfle and Gilson[16] make a more radical suggestion. While usually it is assumed that contraction of the muscle caused by longitudinal tension leads to a bulging of the sarcolemma, they assume that the (twitch) tension acts transversely, causing a bulging of the sarcolemma which leads to contraction. They think that condensation from an elongate to a compact form, at constant volume within the fluid sarcoplasm, may exert

[15] **Bull, H. B.** J.A.C.S **67**: 533. 1945. Cf. also Quart. Bull. Northw. Med. Sch. **20**: 175. 1946.

[16] **Schoepfle, G. M., & A. S. Gilson, Jr.** J. Cell. & Comp. Physiol. **27**: 105–114. 1946. Cf. also the paper by **Gilson, Schoepfle & Walker**, p. 697ff.

lateral pressure and thus indirectly cause a change in longitudinally directed tension. Since the molecular matrix must be inside the filaments studied by Schmitt *et al.*, such a mechanism is somewhat hard to understand. Schoepfle and Gilson[16] investigated the change of elasticity on activation of a muscle by single twitches.[17] They observe actually the differential of a stress-strain curve, *i.e.*, the slopes as a function of tension and of length. As a function of tension, the slopes turn out to be the same for both resting and active muscle, but not as a function of the length. Schoepfle and Gilson interpret their data as a support of their ideas. They also refer to the fact that a stretched, but rigidly fixed, muscle breaks across the belly during the rise of tension following stimulation.

Previous attempts to measure the change of elasticity on activation are obscured by difficulties in the interpretation of the results, as Fenn (*loc. cit.*) points out. Schoepfle and Gilson's work is an improvement, no doubt, but the meaning of their results is not entirely clear at present.

3. Nature of Tension in Active Muscle

What causes the tension in active muscle? The question of activity *versus* recovery energization (*i.e.*, in the older terminology: "chemical *versus* physical theory"), is discussed in detail in the interesting article of Dr. Ramsey* and in other papers of this conference, and in the reviews of Dr. Fenn and Dr. Ramsey referred to before. We shall assume, here, the recovery energization theory. The tension in active muscle is, very likely, due to both internal energy and to entropy. The resting state is the initial state for contraction. Now, as we have seen, the tension in stretched resting muscle is to a large part determined by internal energy. Therefore, at least for contractions up to 20–30 %, the internal energy should play a role, but not necessarily for the largest contractions, *e.g.*, in the delta state. The question of the temperature dependence of the tension-length diagram for active muscle is discussed carefully by Fenn. These data do not give a clear indication of the relative magnitudes of the entropy $Z(S)$ and internal energy $Z(U)$ contributions to the tension. Recent results by Dr. Ramsey on the temperature dependence of tension-length diagrams were analyzed by the method of least squares by Dr. Bull. They seem to be compatible with a pure entropy tension. However, there is a considerable scattering of the original points in this rather difficult experiment. Therefore, in the author's opinion, Dr. Ramsey's ques-

[17] Similar work in tetanus is harder to interpret.
* P. 675ff.

tion (*loc. cit.*), as to whether his data are consistent with an appreciable contribution due to internal energy, can be answered affirmatively.

4. Interpretation of Tension-Length Diagrams

The interpretation of the length-tension diagrams (in particular, those given by Ramsey and Street for single muscle fibers) is ambiguous at present, for, as stated before, we do not know the mechanical coupling between molecular matrix and sarcolemma. Ramsey subtracts the resting tension (due to the sarcolemma) from the total tension and ascribes the difference to the molecular matrix.

This "difference" tension-length curve was considered in the interesting paper by Bull.[15] He ascribed the tension-length behavior to the curling of myosin filaments, and tried to connect the elastic modulus with the molecular weights of the extracted filaments. While this particular interpretation is somewhat problematic at present, Bull's comparison of the resting state of the myosin filaments with stretched and partially crystallized rubber, and of the delta state with amorphous rubber, is probably valid.

Subtraction of the resting tension and consideration of the difference as due to the muscle matrix presuppose that the sarcolemma and the muscle matrix behave as two (non-Hookian) springs in parallel. Since this assumption involves difficulties (*e.g.*, the single twitch experiments of Schoepfle and Gilson and the so-called Weber's paradoxon), it may be safer, at present, to consider, instead, the total tension. We shall assume that the sarcolemma carries the tension, both in the resting and in the active state. (This assumption could be checked experimentally, to a certain extent, by pinching the muscle before it is stimulated, as we suggested before.) The muscle matrix is bound to the sarcolemma by weak bonds before stimulation. After stimulation, the muscle matrix becomes cross-linked with the sarcolemma (vulcanization). These cross-links may be sidewise, or end-to-end, or a mixture of both. Sidewise bonding may lead to a transverse transmission of tension, similar to that suggested by Schoepfle and Gilson. Concurrently with, or perhaps preceding, this cross-linking, a de-cross-linking or destabilization of the actomyosin complex occurs.

On such a picture, Ramsey and Street's tension-length curves of the resting and of the contracting muscle may be considered, in first approximation, as the upper and lower part of the same stress-strain curve, namely, one of the type of FIGURE 15a. The part to the right of ε^* corresponds to the stretching of resting muscle (assuming that the sarcolemma of the resting unstretched muscle corresponds to a stretch

ε^* relative to the maximally contracted state). The part to the left of ε^* corresponds to the contraction of previously unstretched muscle. The length-tension curve of previously stretched active muscle may be qualitatively explained by the decrease of the probabilities of both vulcanization and destabilization by the pre-stretch. However, time effects, too, may lower the total tension in this case.

This description of the length-tension diagram for the contracting muscle is formally very similar to that given by Bull,[15] though the underlying picture is different. The delta state may correspond to amorphous rubber and supercontracted wool. It is possible, however, that, in the maximally contracted state, a still regular folding takes place. Then the analogy with amorphous rubber would be a very rough approximation only, and intermolecular forces may contribute as much as the entropy to the tension.

5. Speed of Shortening and Viscosity

For active muscle, free retraction apparently has not been observed directly, but it may be inferred by extrapolation. Hill's characteristic equation for isotonic contraction,

$$(Z_w + a)v = b(Z - Z_w) \tag{12}$$

(Z_w: weight; v: speed of shortening; Z: isometric tension; a, b: constants), reduces for free retraction to:

$$\frac{a}{b}\, v = Z. \tag{13}$$

If Z is assumed to be constant, then v is also constant and the relation (EQUATION 13) is the equation of motion of a concentrated mass acted upon by a constant force Z and retarded by a viscous medium with a viscous force proportional to the velocity. Originally, Hill assumed that the speed of shortening is determined by such a viscous force. Fenn, however, pointed out that then, for isotonic shortening, the relation,

$$\eta v = (Z - Z_w) \tag{14}$$

(η: coefficient of viscosity), should hold. This equation did not check with the experiments of Fenn and Marsh, who found that muscle lifts heavier weights slower than lighter ones. To explain this fact, Fenn suggested that the speed of shortening is determined by the time needed for energy liberation. Heavier weights need more excess energy liberation, and this takes a longer time. Hill later agreed fundamentally with Fenn and put forward his characteristic EQUATION 12. Hill obtains the

constant, a, also, from measurements of the shortening heat, and thus establishes a connection between speed of shortening and rate of heat production. Hill's formalism is discussed by Fenn (*loc. cit.*) and Ramsey (*loc. cit.*), who point out that Hill's equation was checked only for the initial velocity of shortening. Ramsey assumes a variable velocity of shortening and connects it with the work done on contraction computed from the area of the tension-length diagram. While both Hill's and Ramsey's theories are rather interesting and ingenious, they are phenomenological and do not clarify the nature of the speed of shortening.

In the case of rubber, the coefficient of internal friction decreases with increasing speed of the deformation, if the experiments are evaluated using the concept of a viscous friction proportional to the velocity. In this connection, it would be important to investigate the (free and isotonic) retraction of stretched resting muscle. It is likely that, in this respect, too, resting muscle will show similarity with rubber. By comparison, it will also be possible to see the difference between resting and active muscle, and to decide whether the speed of shortening is intrinsically connected to, and determined mainly by, the chemical processes occurring in the active muscle, or whether the speed of shortening is given mainly by the factors which determine the retraction of stretched rubberlike materials.

We have seen that the speed of free retraction of stretched rubber is determined by the inertia and the internal friction of the particular type of rubber considered. Natural rubber's speed of retraction is a constant and is determined mainly by inertia, while, in the case of Butyl rubber, internal friction slows down the speed as the retraction progresses with time.

6. Thermoelastic Effect of Active Muscle

Azuma found (in Hill's laboratory) that active muscle, on stretching, exhibits a cooling effect like a normal solid, in contrast to the rise in temperature on stretching of rubberlike materials and of resting muscle (below the inversion point).

We have pointed out, in SECTION A 12, that rubber vulcanized in a stretched state shows a normal thermoelastic behavior. If Azuma's effect is a thermodynamic phenomenon (*i.e.*, if it is not a time-dependent process), and if the "vulcanization" of the muscle on stimulation is more than an analogy, then it may be explained in the suggested way. At any rate, further work along these lines would be interesting.

There seems to be no analogy with rubber for the interesting effects of latency relaxation studied in such a remarkable manner by Dr. Sandow.*

APPENDIX

1. Statistical Treatment of the Bundle Model

The quantitative treatment of the rubber chain is again very much the same as that used for a gas. For simplicity, an ideal gas may be employed. A configuration of gas molecules is the more probable, the greater the volume occupied by the gas molecules. Hence, it is plausible to put this configuration (or thermodynamic) probability, C, proportional to the volume, V:

$$C(V) = AV; A = \text{const.} \tag{1a}$$

In order to derive an equation of state, we have to connect the configurational probability $C(V)$ with the entropy $S(V, T)$ of the gas. Obviously, nature will prefer more probable configurations, as compared to less probable ones. This tendency of nature requires that the entropy changes in conformance with the configurational probability, a statement qualitatively equivalent to the second law of thermodynamics. Quantitatively, the connection between $S(V, T)$ and $C(V)$ is given per molecule by Boltzmann's principle:

$$s(V, T) = k \log C(V) = k \log(A \cdot V), \tag{2a}$$

and per mole with N molecules introducing the gas constant $R = Nk$,

$$S = Ns = R \log (A \cdot V). \tag{3a}$$

The equation of state $p = p(V, T)$ follows from the entropy by purely thermodynamical considerations:

$$p = T\left(\frac{\partial S}{\partial V}\right)_T = \frac{RT}{V}. \tag{4a}$$

For a rubber chain containing N freely rotatable links, the configurational probability $C(L)\,dL$ will be a measure of the probability that the ends of the chain will be apart by a distance between L and $L + dL$. It is natural to expect that $C(L)$ will have the form of the error law, *i.e.*, of the Gaussian distribution:

* Cf. pp. 895–930.

$$C(L)dL = Be^{-\beta L^2}dL;\ B = \text{const.}$$

Here:

$$\beta = \frac{3}{2}\cdot\frac{1}{Nl_\delta^2};\ l_\delta = l\tan\frac{\delta}{2} \qquad (5a)$$

with $l = 1.54$ Å, the distance of the $C - C$ bond, and $\delta = 109°$, the valence angle. Rigorous analysis, which we omit here, verifies this expectation.

EQUATION 5a corresponds to EQUATION 1a. To the volume, V, of a gas, there corresponds the length, L, between the ends of the rubber chain.

Again, Boltzmann's principle will connect the configurational probability $C(L)$ with the entropy $s(L, T)$ of the rubber chain:

$$s(L, T) = k\log C(L) = k\log B - k\beta L^2, \qquad (6a)$$

and for a *bundle* of M parallel chains:

$$S(L, T) = M\cdot s(L, T). \qquad (7a)$$

The equation of state $Z = Z(L, T)$ where Z designates the force stretching the rubber normalized to unit original cross-section, follows from the entropy by the analogue of EQUATION 4a:

$$= -T\left(\frac{\partial S}{\partial L}\right)_T = M\cdot 2\beta\cdot kT\cdot L = KTL. \qquad (8a)$$

Applied to bulk rubber, L may be interpreted as the relative length, *i.e.*, the extended length divided by the original length.

This equation explains *qualitatively* the three thermoelastic effects exhibited by rubber, and agrees fairly well with experiment for medium extensions.

A. For the rise in temperature on fast stretching, thermodynamics yields the formula:

$$\Delta T = \frac{T}{C_L}\int_1^L\left(\frac{\partial Z}{\partial T}\right)_L dL. \qquad (9a)$$

Inserting the value of $(\partial Z/\partial T)_L$ obtained from EQUATION 8a yields:

$$\Delta T = \frac{T}{C_L}\cdot\frac{K}{2}\cdot(L^2 - 1). \qquad (10a)$$

This formula explains the experimental results above the thermoelastic inversion point given in FIGURE 1, but it fails both for very low and very high extensions.

B. The length of the rubber sample as a function of temperature, when it is stretched by a constant load, may be characterized by the linear thermal expansion coefficient parallel to the direction of the stress.

$$\alpha_{\parallel} = \frac{1}{L}\left(\frac{\partial L}{\partial T}\right)_Z \cdot \tag{11a}$$

$\alpha_{\parallel}$ may be computed from EQUATION 8a directly, or using the thermodynamic relation:

$$\left(\frac{\partial L}{\partial T}\right)_Z = -\left(\frac{\partial L}{\partial Z}\right)_T \cdot \left(\frac{\partial Z}{\partial T}\right)_L \cdot \tag{12a}$$

One obtains:

$$\alpha_{\parallel} = -\frac{1}{T}; \tag{13a}$$

i.e., the linear thermal expansion coefficient of stretched rubber has exactly the same magnitude [in the approximation of EQUATION 8a] as the cubic expansion coefficient of an ideal gas, but it has a negative sign. Thus, EQUATION 13a explains why rubber stretched isotonically contracts visibly when heated. $\alpha_{\parallel}$ (in the approximation of EQUATION 8a) is quite unrelated to the linear expansion coefficient of *unstretched* rubber, which is of the order of $2 \cdot 10^{-4}$. Again, EQUATION 13a only holds well above the thermoelastic inversion point.

C. For a sample of rubber held stretched to a constant length, EQUATION 8a gives the (proportional) increase of stress with temperature.

The upward curvature of the stress-strain curve, as shown in FIGURE 5, may also be explained on the basis of the bundle model. For it can be proved that the Gaussian distribution (EQUATION 5a) holds only if the "length" L between the ends of the chain is small, compared to the maximum length which the chain has when completely extended.

A more exact treatment of a one-dimensional model yields:

$$Z = M\frac{kT}{l}f(t) \tag{14a}$$

where $$t = \frac{L}{Nl} \text{ and } f(t) = \tanh^{-1} t = 1/2 \log \frac{1+t}{1-t}.$$

The function $t = \tanh f$ occurs in Ehrenfest's theory of magnetism. The function $f(t)$ is shown as curve II in FIGURE 20. This curve explains the upward curvature of the rubber stress-strain curves simply as due to deviations from the Gaussian law, as the chains are straightened out more and more by the stress.

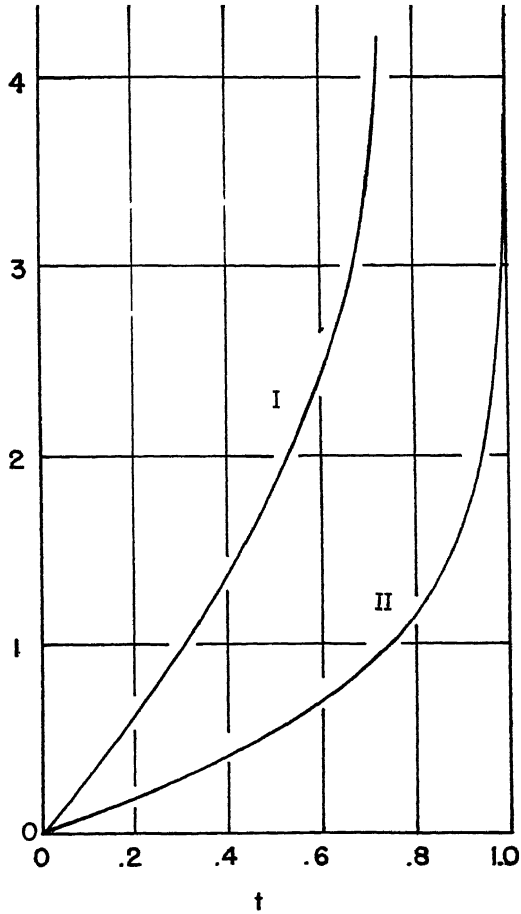


FIGURE 20. Inverse Langevin function (I) and inverse tanh function (II).

A three-dimensional model corresponding to the one-dimensional model just treated yields:

$$f(t) = \mathfrak{L}^{-1}(t); t = \mathfrak{L}(f) = \coth f - 1/f \qquad (15a)$$

where $L(f)$ is the Langevin function occurring in Langevin's theory of paramagnetism and Debye's theory of polar molecules.

2. Equation of State for Incompressible Rubber Consisting of Gaussian Chains

We consider the irregular cross-bonded network as described above. This network can be replaced by a regular network of cubic symmetry

of independent chains without any loss of mathematical rigor. The chains of this regular network may run parallel to the three coordinate axes. For this network, it is easy to obtain the equation of state for stretched rubber.

We make two fundamental assumptions:

A. The Gaussian distribution is valid for $C(L)$, the configurational probability: *i.e.*, the rubber chains are extended only to a length small compared to the maximum length.

B. Rubber is incompressible: *i.e.*, it does not change its volume when stretched.

Assumption B holds experimentally for soft gum compounds up to about 300 % extension, according to Holt and McPherson.

We consider a unit cube of unstretched rubber. After a force Z is applied in the Z-direction, the dimensions of the resulting paralleliped will be L_x, L_y, L_z. For unilateral deformations in the Z-direction we have the relation:

$$L_x = L_y. \tag{16a}$$

According to our Assumption B and EQUATION 16a,

$$V = L_x L_y L_z = L_y^2 L_z = 1. \tag{17a}$$

As described earlier, there will be three forces acting in our model:

Z_1, the inward pull due to the contracting force;

Z_2, the outward push due to the internal pressure of the rubber liquid*;

Z_3, the outward pull due to the external force.

Because of the equivalence of the irregular and of the regular cubic network, we may use for the contracting force EQUATION 8a, derived for a bundle of M parallel chains.

Then,

$$Z_1 = KTL_x. \tag{18a}$$

The internal pressure, P, may be calculated from the condition of the equilibrium of the forces applied to one of the sides (L_y L_z) of the cube upon which no external force acts. The outward push of this pressure is balanced by the pull of M molecular chains, each extended to length L_x and thus exerting an average force of $Z_1 = KTL_x$.

Then,

$$PL_yL_z = KTL_x; \tag{19a}$$

using EQUATION 16a, we obtain:

* This "liquid" is fictitious as far as it simulates effects of steric forces and molecular bombardments, but more nearly real as far as it represents a sol component.

$$P = KT/L_x. \tag{20a}$$

Thus, the internal pressure of rubber decreases for tension and increases for compression, in accordance with physical expectations.

Considering the forces acting on the end faces (L_xL_y), we have:

$$\begin{aligned} Z_1 &= KTL_x \\ Z_2 &= PL_yL_z = KT/L_x^2. \end{aligned} \tag{21a}$$

For equilibrium between the three forces, the sum of the outward forces must be equal to the inward pull, or

$$Z_2 + Z_3 = Z_1. \tag{22a}$$

Inserting the value of Z_1 and Z_2, we obtain:

$$Z = Z_3 = Z_1 - Z_2 = KT[L_x - 1/L_x^2]. \tag{23a}$$

This is the stress-strain relation we were looking for. Physically, it states that the external force balances the difference between the tension due to the internal Brownian motion and that due to the internal pressure.

The stress-strain relation (EQUATION 23a) holds both for extension ($L_x > 1$) and for compression ($L_x < 1$), thus removing an outstanding difficulty of the bundle model. It holds also, of course, for small extensions. At large L_x, EQUATION 23a has an asymptote which passes through the origin. At small L_x, EQUATION 23a has the stress axis as its asymptote.

FIGURE 21, curve B, is an experimental stress-strain curve obtained at Notre Dame. The circles are theoretical points, as given by EQUATION 23a.

Young's modulus, E, may be defined as the slope of the stress-strain curve. From EQUATION 23a we obtain:

$$E(L_x) = dZ/dL_x = KT[1 + 2/L_x^3]. \tag{24a}$$

For $L_x = 1$, E has the value:

$$E(1) = 3KT, \tag{25a}$$

which is three times the value:

$$E(L_x \gg 1) = KT. \tag{26a}$$

To see the range of validity of Hooke's Law, we introduce, instead of the relative length, L, the fractional extension or strain $\epsilon = L - 1$. EQUATION 23a becomes:

$$Z = KT\left[(1 + \epsilon) - \frac{1}{(1 + \epsilon)^2}\right] = 3KT\epsilon[1 - \epsilon + 4/3\epsilon^2 \text{---}] \tag{27a}$$

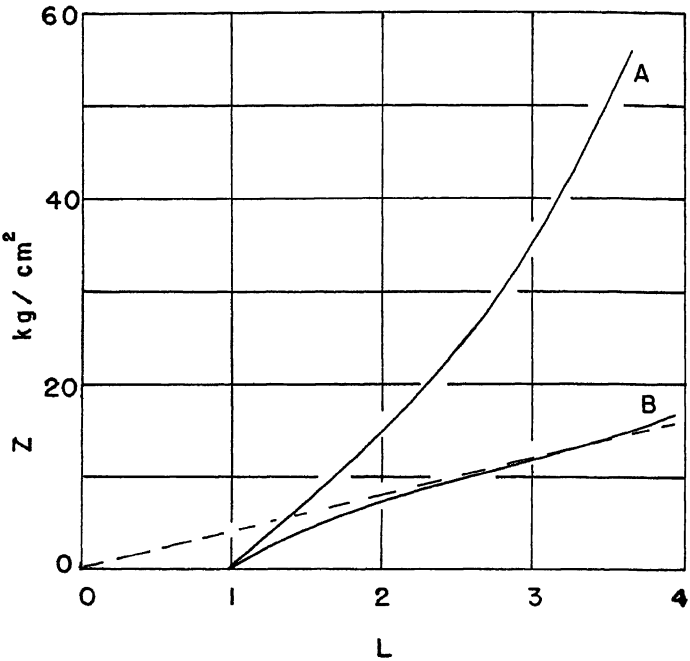


FIGURE 21 Curve A True stress *vs* relative length for an accelerated pure gum stock Curve B Nominal stress *vs* relative length corresponding to Curve A The dashed line represents the asymptote of the theoretical curve

From EQUATION 27a, it follows that, for extensions as small as 5 %, there is a 5 % deviation from Hooke's Law. Young's modulus is, of course, the same for both variables L and ε.

3. The Thermoelastic Effects for the Network Model

For the rise in temperature on fast stretching, insertion of EQUATION 23a into EQUATION 9a yields:

$$\Delta T = \frac{KT}{2C_L}(L_z^2 + L_z - 2)\left(\frac{L_z - 1}{L_z}\right). \tag{28a}$$

The linear thermal expansion coefficient parallel to the direction of the stress, $\alpha_{\|}$, takes on the form:

$$\alpha_{\|} = 1/L_z\left(\frac{\partial L_z}{\partial T}\right)_z = -\frac{1}{T}\cdot\frac{L_z^3 - 1}{L_z^3 + 2} \tag{29a}$$

This expression does not differ significantly from EQUATION 13a, to which it reduces for $L_z \gg 1$. A novel feature, however, is the possi-

bility of computing the linear thermal expansion coefficient perpendicular to the direction of the stress:

$$\alpha_\perp = \frac{1}{L_y}\left(\frac{\partial L_y}{\partial T}\right)_Z = \frac{1}{2T} \cdot \frac{L_z^3 - 1}{L_z^3 + 2}. \tag{30a}$$

In deriving EQUATION 30a, the constancy of the volume of rubber during stretching, EQUATION 17a, was used. The possibility of introducing $\alpha_\perp$ rests on the three-dimensional character of the new network model in contrast to the essentially one-dimensional bundle model.

EQUATIONS 29a and 30a do not contain the cubic expansion coefficient of unstretched rubber and, consequently, do not show a thermoelastic inversion point.

4. Generalization of the Stress-Strain Relation to Larger Extensions

The generalization of our model to larger extensions involves three parts:

A. The contracting force for a single chain must be generalized to hold beyond the range of validity of the Gaussian distribution.

B. An irregular network must be considered and reduced to a form in which the expression for the contracting force for a single chain may be utilized for the contracting force of the whole network.

C. The contracting force of the network must be balanced by the internal pressure of the rubber liquid and any external forces.

The result for Part A, the contracting force for a single chain, was recorded earlier. It has not as yet been possible to treat Part B in the same rigorous fashion as that employed in the case of Gaussian chains. Instead of carrying out the reduction from an irregular to a regular network, we simply assume that, in the unstretched state of the material, each independent chain has an extension equal to κ times its maximum length. The maximum stretch is thus given by a factor $1/\kappa$, a quantity which will depend upon the constitution of the particular rubber considered. κ is an adjustable parameter in our treatment.

Considering again the equilibrium of forces, and using for the contracting force the three-dimensional result (EQUATION 15a) we obtain:

$$Z = K'T\left[\mathfrak{L}^{-1}(L_z\kappa) - \frac{3\kappa}{L_z^2}\right]. \tag{31a}$$

This relation involves essentially two parameters, the scale factor K'

and the new maximum stretch parameter κ. When the extension is small compared to the maximum,

$$\mathfrak{L}^{-1}(t) \cong 3t \tag{32a}$$

$$Z \cong 3\kappa K'T\left[L_z - \frac{1}{L_z^2}\right]; \tag{33a}$$

i.e., EQUATION 31a reduces to the Gaussian expression (EQUATION 23a), with $3\kappa K'$ taking the place of the old parameter K.

EQUATION 31a yields an S-shaped stress-strain curve. The occurrence of the inflexion point, *i.e.* of the S-shape, is explained as due to the superposition of the effects of the contracting force and those of the internal pressure.

Referred to the actual cross-section, the nominal stress-strain relation (EQUATION 31a) takes on the form:

$$Z' = K'T[L_z\mathfrak{L}^{-1}(L_z\kappa) - 3\kappa/L_z]. \tag{34a}$$

This true stress-strain curve does not show the S-shape of the nominal stress-strain curve. FIGURE 6, curve A, shows the true stress-strain curve. Curve B is the nominal stress-strain curve. EQUATION 33a explains the S-shape of curve B quite well, and curve A is well described by EQUATION 34a.

5. Influence of the Thermal Expansion on the Stress-Strain Relation

With the successive refinements introduced up to this point, all of the peculiar thermoelastic properties of rubber may be explained, with the exception of the thermoelastic inversion point. To explain this phenomenon, we must take into account the cubic thermal expansion of unstretched rubber, characterized by a coefficient of thermal expansion, α, defined by the equation:

$$V(T) = V(T_o)[1 + \alpha(T - T_o)]. \tag{35a}$$

The previous stress-strain relation (EQUATION 23a) yields zero stress for $L = 1$, *i.e.*, the unstretched state. When thermal expansion is considered, the unstretched state is characterized by $L_z = 1 + \alpha/3(T - T_o)$. The previous EQUATION 23a should be changed to:

$$Z = KT\left[L_z - \frac{1 + \alpha(T - T_o)}{L_z}\right], \tag{36a}$$

which gives:

$Z = 0$, when $L_z = 1 + \alpha/3(T - T_o)$. The thermal expansion enters only into the internal pressure term, and the contracting force (the

first term in the bracket) is independent of α. We emphasize that L_z, by definition, is independent of α, the original length being taken at $T = T_o$.

EQUATION 36a predicts a thermoelastic inversion point as the solution of the equation:

$$\left(\frac{\partial Z}{\partial T}\right)_{L_z} = K\left[L_z - \frac{1 + \alpha(2T - T_o)}{L_z^2}\right]. \tag{37a}$$

Thus, the critical extension at which the stress-temperature curves, the isometrics, change their slope is given by the isometric thermoelastic inversion point:

$$L_z^3 = 1 + \alpha(2T - T_o) = 1 + \alpha T_o \text{ for } T = T_o. \tag{38a}$$

Because of the thermodynamic relation,

$$\left(\frac{\partial L_z}{\partial T}\right)_Z = -\left(\frac{\partial L_z}{\partial Z}\right)_T \left(\frac{\partial Z}{\partial T}\right)_{L_z}, \tag{39a}$$

the isotonic thermoelastic inversion point is identical with the isometric one.

From EQUATIONS 33a and 35a, the linear thermal expansion coefficients in the direction of the stress and perpendicular to it, $\alpha_\parallel$ and $\alpha_\perp$, may be computed. One obtains for $T = T_o$:

$$\alpha_\parallel(L_z) = \frac{1}{L_z}\left(\frac{\partial L_z}{\partial T}\right)_Z = \left(\frac{-1}{T}\right)\frac{L_z^3 - 1}{L_z^3 + 2} + \frac{\alpha}{L_z^3 + 2}, \tag{40a}$$

and:

$$\alpha_\perp(L_z) = 1/L_y\left(\frac{\partial L_y}{\partial T}\right)_Z = +(1/2T)\cdot\frac{L_z^3 - 1}{L_z^3 + 2} + \frac{\alpha L_z^3 + 1}{2L_z^3 + 2}. \tag{41a}$$

In deriving EQUATION 36a, the constancy of the volume of rubber during stretching was used, namely:

$$V = L_y^2 L_z = 1 + \alpha(T - T_o). \tag{42a}$$

In the unstretched state $L_z = 1$, we have:

$$\alpha_\parallel(1) = \alpha_\perp(1) = \alpha/3, \tag{43a}$$

as for any isotropic material. $\alpha_\parallel$ changes from positive to negative values with increasing L_z, whereas $\alpha_\perp$ is always positive.

The formula corresponding to EQUATION 10 becomes for $T = T_o$:

$$\Delta T = \frac{KT}{2C_L}\left[L_z^2 + L_z - 2(1 + \alpha T)\right]\left[\frac{L_z - 1}{L_z}\right]. \tag{44a}$$

According to this formula, $\triangle T$ starts with the value zero at $L_z = 1$, reaches a minimum at the thermoelastic inversion point as given by EQUATION 38a, then becomes zero for the adiabatic thermoelastic inversion point, determined by:

$$L_z^2 + L_z - 2(1 + \alpha T_o) = 0, \tag{45a}$$

and takes on positive values for higher values of L_z.

The introduction of the cubic thermal expansion of rubber into the stress-strain relation (EQUATION 31a) is simple, because only the internal pressure is modified. One obtains:

$$Z = K'T\left\{\mathfrak{L}^{-1}(L_z\kappa) - \frac{3\kappa[1 + \alpha(T - T_o)]}{L_z^2}\right\}. \tag{46a}$$

6. Comparison of the Theoretical Stress-Strain Relations with Experiment

In comparing the theory with experiment, a resolution of the stress into its additive parts due to entropy and internal energy has to be carried out, both for the experimental results and for the theory.

In van der Waals' equation of a real gas,

$$p = -a/V^2 + RT/V - b, \tag{47a}$$

the first term, $p(U) = -a/V^2$, is due to the internal energy U, and the second term, $p(S) = RT/V - b$, is due to the entropy S. According to thermodynamics:

$$p(U) = -\left(\frac{\partial U}{\partial V}\right)_T, \; p(S) = T\left(\frac{\partial S}{\partial V}\right)_T = +T\left(\frac{\partial p}{\partial T}\right)_V, \tag{48a}$$

so that:

$$p = p(U) + p(S) = -\left(\frac{\partial U}{\partial V}\right)_T + T\left(\frac{\partial p}{\partial T}\right)_V. \tag{49a}$$

For a unilateral stress Z, instead of a uniform pressure p, we must replace p by $-Z$ (since pressure and stress act in opposite directions), and the volume V by the relative length L. Then we obtain:

$$Z = Z(U) + Z(S) = \left(\frac{\partial U}{\partial L}\right)_T + T\left(\frac{\partial Z}{\partial T}\right)_L. \tag{50a}$$

Using EQUATION 50a in conjunction with experimental stress-temperature data, we may calculate $Z(U)$ and $Z(S)$ and plot them as functions of L. Using EQUATIONS 50a, 36a and 46a, we obtain as theoretical expressions for $Z(U)$ and $Z(S)$:

$$Z(U) = \frac{\alpha K T^2}{L_z^2}\,;\; Z(S) = KT\left\{L_z - \frac{[1 + \alpha(2T - T_o)]}{L_z^2}\right\} \tag{51a}$$

$$Z(U) = \frac{3\kappa K'\alpha T^2}{L_z^2}\,;\; Z(S) = K'T\left\{\mathfrak{L}^{-1}(L_z\kappa) - \frac{3\kappa[1 + \alpha(2T - T_o)]}{L_z^2}\right\}. \tag{52a}$$

For further discussion of the theory and comparisons with experimental data, the reader should refer to an article by James and Guth (J. Chem. Phys. 11: 455. 1943). It should be emphasized, however, that, in the case of most synthetic rubbers, the agreement between theory and experiment is not as good as it is for natural rubber.

7. General Theory of Time Elasticity

Time elasticity shall designate any process in which stresses cause deformations with the time element as an important factor. A general formalism may be given in agreement with the factorization (EQUATION 10) by a generalization of Boltzmann's theory of time elastic processes.

The fundamental integral equation is ($\varepsilon = L - 1 =$ strain):

$$Z(t) = F[\epsilon(t)] - \beta\int_{-\infty}^{t} \phi\,(t - t')F[\epsilon(t')]dt'. \tag{53a}$$

β is a (small) constant. $\beta\phi(t - t')$ is the memory function of Boltzmann. EQUATION 53a reduces to Boltzmann's equation, if the instantaneous stress-strain relation,

$$Z = F(\epsilon), \tag{54a}$$

reduces to Hooke's Law,

$$Z = E\cdot\epsilon, \tag{55a}$$

where the constant E denotes Young's modulus.

For relaxation:

$$\epsilon(t) = \epsilon(t') = \begin{cases} 0, & t<0 \\ \epsilon = \text{const.}, & t>0. \end{cases}$$

EQUATION 53a reduces to

$$Z(t, \epsilon) = F(\epsilon)[1 - \beta\Psi(t)] = F(\epsilon)G(t), \tag{56a}$$

where

$$\Psi(t) = \int_o^t \phi\,(t - t')dt'. \tag{57a}$$

EQUATION 56a is identical with EQUATION 10 of SECTION B1. (For simplicity, we did not indicate, in EQUATION 53a, the temperature as another variable.)

For creep, we have to invert our fundamental EQUATION 53a. For short times, this is simple. Then the term with β will be small, and terms of order β^2 or higher may be neglected. We obtain:

$$\epsilon(t) = F^{-1}\left\{Z(t) + \beta\int_{-\infty}^{t}\phi(t-t')Z(t')dt'\right\}, \tag{58a}$$

and for creep:

$$Z(t) = Z(t') = \begin{cases}0, & t<0\\ Z = \text{const.}, & t>0,\end{cases}$$

$$\epsilon(t) = F^{-1}\left\{Z[1 + \beta\Psi(t)]\right\}. \tag{59a}$$

EQUATION 59a yields factorization for creep in the special case of Hooke's Law, where $\varepsilon = F^{-1}(Z) = Z/E$, but not in the general case of EQUATION 54a. We notice that, had we assumed factorization for creep, we would have arrived at a formula different from EQUATION 58a, which would have contradicted EQUATION 53a. Our experiments show that, in contrast to relaxation, factorization does not hold for creep.

All time-elastic phenomena may be derived from the fundamental EQUATION 53a. For instance, we can obtain the dependence upon time of stress-strain curves obtained at constant rate of loading or of stretching. The fast processes of forced or free vibrations may be treated, using the same equation as a general frame.

To determine the memory function, one needs a molecular model. The memory function may arise from different mechanisms in different regions of frequency. The memory function may be expressed by a distribution function for "relaxation times."

8. Theory of the Retraction of Stretched Rubber

We consider a uniform elastic rod of length l_0, fixed at one end, with linear density ρ in the unstretched state. We assume that Hooke's Law holds, with Young's modulus E, and that frictional forces are proportional to the rate of deformation. Let x be a measure of distance in the material of the unstretched rod, starting at the fixed end, and let $u(x)$ be the displacement of the material corresponding to the indicated value of x. In the absence of body forces, the behavior of the rod in extension and retraction is governed by the wave equation:

$$\rho\frac{\partial^2 u}{\partial t^2} = f\frac{\partial^3 u}{\partial x^2\partial t} + E\frac{\partial^2 u}{\partial x^2}. \tag{60a}$$

If the rod, initially, is uniformly stretched with fractional extension ϵ, and is released at time $t = 0$, the boundary conditions are:

$$\partial u/\partial x = \epsilon,\ \partial u/\partial t = 0 \text{ for all } x \text{ at } t = 0,$$
$$u = 0 \text{ for all } t \text{ at } x = 0, \tag{61a}$$
$$f\frac{\partial^2 u}{\partial x \partial t} + E\frac{\partial u}{\partial x} = 0 \text{ for } x = l_0 \text{ when } t > 0.$$

In the limiting case of small friction, $f = 0$, the solution becomes, with $v = (E/\rho)^{1/2}$:

$$\begin{aligned} u(x, t) &= \epsilon x & l_0 - x > vt, \\ u(x, t) &= \epsilon(l_0 - vt) & l_0 - x < vt < 1. \end{aligned} \tag{62a}$$

According to this solution, a pulse of acceleration arises at the free end at the moment it is released, and travels, with respect to the unstretched material, at a rate v; its rate of transmission in space of $v_p = (\epsilon + 1)v$. Each portion of the bar is stationary until it is reached by this pulse. The velocity of the tip of the bar is $v_t = -\partial u/\partial t = \epsilon v$. As the pulse passes over it, the stress is completely relaxed, and the material reaches its final velocity ϵv. When the acceleration pulse reaches the fixed end of the material, after time $t_c = l_0/v$, the bar is unstretched and in uniform motion at the rate ϵv, the original potential energy is completely converted into kinetic energy. After this time, a solution other than EQUATION 62a begins to apply. The wave pulse is reflected at the fixed end and a pulse of deceleration begins to progress back down the bar. In the case of a rod in which transverse motion is not prevented, as in the case of a retracting rubber band, buckling begins at the moment of reflection of the pulse.

The first element of the rod has a very small mass. Applying a finite tension, a very large acceleration results, and the final velocity is attained instantly. This is in contrast to the free vibration of a concentrated mass, which always starts with zero velocity. For the actual stylus experiments, the velocity starts with zero. This is caused by the weight of the stylus and of the unstretched rubber at the free end of the sample.

Going back to EQUATION 60a, let us consider the case involving large internal friction. Here, the inertial term can be neglected and the solution becomes:

$$u(x,t) \cong \epsilon x e^{-(E/f)t} \tag{63a}$$

In this case, we neglect the wave pulse and, hence, this equation holds

only after the wave pulse has passed down the sample. Here we see that, at any given time, the displacement (and also the velocity) of any point on the sample is proportional to its distance from the fixed end. Also, of course, the sample approaches the unstretched state exponentially with time.

For Butyl rubber, the first part of the elongation-time curve is mainly determined by inertia, while, in the last part, internal friction predominates.

MEMBRANE CHANGES DURING EXCITATION AND INHIBITION OF THE CONTRACTILE MECHANISM

By S. W. Kuffler*

Department of Physiology, University of Chicago, Chicago, Illinois

In his introduction, Dr. Sandow referred to our ignorance of the chain of events which connect the muscle membrane and the contractile elements. The aim of this discussion is to point out at least one approach to this question, namely, the relationship between membrane changes and initiation and maintenance of contraction.

There has existed, for a long time, ample proof for propagated muscle impulses, and this fact has formed the basis for much experimental work (*e.g.*, Young, 1941). More recently, the conditions of excitation and propagation have been worked out in detail on completely isolated single muscle fibers (Kuffler, 1942 a & b). Since several workers still hold a contrary view, it must be stressed that there seems to be no *essential* difference in excitation of nerve or muscle membranes. The potential changes which lead up to the initiation of impulses, or which accompany propagated muscle responses, are similar to those observed in nerve (Katz, 1939).

A. CONTRACTURE IN VERTEBRATE SKELETAL MUSCLE

The term, *contracture,* has a variety of meanings, and is used for many different states or processes in muscles. For our present purpose, contracture may be defined as a prolonged, reversible, but not propagated shortening of the muscle (Gasser, 1930).

There are several advantages in studying contractures, as compared with the usual twitch-response. The latter cannot be controlled, the muscle impulse takes care of the whole cycle and, as it passes along, initiates the contraction along the entire fiber. Contractures, however, can be regulated, their strength can be graded, the "activation" can be stopped at will, and the visible changes, being slow, can easily be observed under a microscope. Contractures are usually obtained by direct excitation, not through the nerve, although, under certain conditions, local shortening at the neuromuscular junction may follow nerve stimulation (Bremer, 1932; Cowan, 1940; Feng, 1941). Con-

* Seymour Coman Fellow in the Department of Physiology.

tractures have been set up in three ways: (i) by constant currents; (ii) by drugs which produce contractures directly on application; (iii) by treatment of muscles with veratrine, which exerts its effect after a muscle impulse has been set up.

Electric recordings from whole muscles showing contractures are frequently difficult to interpret. Whenever twitches and contractures occur simultaneously, but in different fibers, the propagated impulses may tend to obscure the relatively small potentials which accompany contractures. Further, if contractures are confined to discrete parts of an otherwise inactive muscle, their potentials may easily be overlooked. It is, therefore, essential to record accurately from the site of contractures. The completely isolated muscle fiber suspended in paraffin oil is most suitable for this purpose, as it yields relatively large potentials.

1. Veratrine

In concentrations as used in a recent study (Kuffler, 1945), veratrine does not affect the resting potential or the electric time constant of the muscle membrane. A drop of the drug is placed on a small area (1–2 mm.) on an isolated muscle fiber suspended horizontally in paraffin oil. A muscle impulse then can be set up somewhere along the fiber course, and the potential changes under the veratrine electrode are recorded, relative to the normal, untreated, muscle part. Passing through a veratrinized area, a muscle impulse leaves behind the well-known, prolonged afterpotential. When this reaches a critical magnitude, it will set up a further impulse which, in turn, initiates another one. In this way, a tetanus results, originating at the veratrinized region (Kuffler, 1945; also, Feng, 1938). The frequency and number of impulses depend on the condition of the muscle and on the drug concentration. Similarly, the development of contractures which generally follow the impulses is determined by the same factors. In PLATE 2, FIGURE 1A, two consecutive recordings are illustrated. Following the initial, electrically produced impulse, a tetanus arises at the veratrine-electrode, reaching a frequency of well over 250 per sec. in PLATE 2, FIGURE 1A(b). The propagated impulses gradually decrease in height and, eventually, a transition is seen to small oscillatory potentials which in turn develop into a smooth maintained potential. During all the recordings of PLATE 2, FIGURE 1, a contracture was observed confined to the veratrine-treated region.

The following question arises: Are the potential changes, which persist while the contracture lasts, caused by the activity in the contrac-

tile system, or do they give rise to the contracture in a similar way as the propagated muscle impulse to the normal twitch-like contraction? The latter mechanism is suggested for the following reasons:

a. Similar potential changes, obviously originating in the membrane, have been observed in nerve treated with veratrine (Graham and Gasser, 1931; Acheson and Rosenblueth, 1941).

b. In the drug-treated area, there is a gradual transition from propagated muscle impulses which diminish in size, to the oscillatory potential changes which seem to be "abortive," not fully propagated responses, as can be well seen in PLATE 2, FIGURE 1A(b). The latter have previously been observed in fatigued or partially refractory muscle (Kuffler, 1942 a) and during curarization (Eccles and O'Connor, 1941). These abortive impulses arise at a similar threshold as the previously propagated impulses, and again a gradual transition occurs, now to the smooth maintained negative potential. It is suggestive that the same process which causes the initial membrane depolarization, leading up to propagated responses and to the associated contraction, also causes the subsequent potentials giving rise to partially and non-propagated contraction, *i.e.*, contracture.

c. There is a definite relationship between the potential size and the threshold for muscle shortening. This threshold, however, does not appear to be as critical as for propagated responses, since it is difficult to detect the first visible sign of contractures (see also SECTION B).

d. A similar sequence of events has been observed at the cathode of a prolonged current pulse. There is little doubt that, in this case, the muscle contraction arises subsequently to the membrane changes which appear immediately on current application (Kuffler, 1942 b). The "cathodal contracture" lasts for the approximate duration of the applied current which causes the membrane negativity during its flow.

2. Contractures Set Up by Drug Application

Unlike veratrine, drugs like acetylcholine (ACh), nicotine, caffeine or KCl cause membrane changes directly on application. While the latter is effective anywhere on the muscle, the other drugs show a selective excitatory action on the endplate regions (Kuffler, 1943). PLATE 2, FIGURE 2a illustrates potentials recorded from the neuromuscular junction while ACh was applied. A relatively slow depolarization, caused by the drug, sets up at a critical threshold propagated impulses. Again, there is a transition to a few "abortive" responses and to the maintained negative potential associated with contracture. The

same concentration of ACh, if applied to an endplate-free portion of the muscle, would be ineffective.

In experiments in which contractures were set up by direct drug application or by constant current pulses, it was convenient to soak the muscles first in novocaine solutions (0.1 to 0.5 %). These concentrations did not reduce the resting potential and did not change the electric time constant of the membrane as tested by constant currents. At the same time, novocaine prevents the setting-up of propagated impulses and contractions, while the thresholds at which contractures first appeared did not seem to change appreciably (Kuffler, 1946).

3. Inhibition of Contractures by Applied Currents

The inhibitory effect of anodal currents is well known in nerve and muscle physiology. Biedermann (1895) reported that veratrine contracture could be inhibited at the anode of a constant current. Bremer (1932) and Gelfan (1933) caused relaxation of contractures which followed indirect or direct muscle stimulation.

If prolonged contractures are set up by drug application (PLATE 2, FIGURE 2), an applied anodal current relaxes the contracture, which then redevelops on withdrawal of the current. Conversely, cathodal currents will augment the drug-contracture (Kuffler, 1946).

These experiments again suggest strongly the role of the membrane in initiating and maintaining contractures. Apparently, the anodal current tends to restore the membrane to its original polarized state, thereby removing the stimulus which had caused the contracture. On cessation of the repolarization at the anode, the original action of the drug (KCl or ACh in PLATE 2, FIGURE 2), which had been "neutralized" during the current flow, reasserts itself by depolarizing the membrane again. This then causes a return of the contracture.

B. NEUROMUSCULAR EXCITATION AND INHIBITION

The neuromuscular system of crustaceans offers some striking advantages for the study of excitation and inhibition of the contractile elements. (i) Local contractions occur in crustaceans, and a great deal of the normal activity seems to proceed without propagated muscle responses. Further, the local contractions can be accurately controlled by varying the frequency and number of nerve impulses. (ii) Crustacean muscles are supplied by antagonistic nerve fibers having an excitatory and inhibitory effect (Harreveld and Wiersma, 1939).

Only a brief account on neuromuscular excitation will be given here. For details see Katz and Kuffler, 1946.

1. Excitation

In crustacean, as in vertebrate, muscles, the motor nerve impulse sets up an endplate potential (e.p.p.) at the neuromuscular junction. In normal vertebrate muscle, the excitatory role of the e.p.p. consists in setting up the propagated muscle impulse, which then causes the propagated contraction. E.p.p.'s alone in curarized muscles (Eccles, Katz, and Kuffler, 1941) do not excite the contractile elements. Under certain conditions, however, neuromuscular contractions have been observed (Cowan, 1940; Feng, 1941), especially in eserinized preparations where the e.p.p. is large and prolonged (Eccles, Katz, and Kuffler, 1942).

The electrical responses of crustacean muscles resemble partially curarized vertebrate preparations. In crustaceans, however, distinct mechanical responses accompany a series of e.p.p.'s. A definite correlation has been found between the activation of the contractile elements and the size of the e.p.p.'s. The size of *individual* e.p.p.'s varies greatly in different muscles. Further, with repetitive stimulation, *successive* e.p.p.'s in many muscles continue to grow in amplitude. The degree and time course of this "facilitation" also vary greatly. Depending upon the initial size and rate of growth of successive e.p.p.'s, "fast" and "slow" systems can be distinguished (Katz, and Kuffler, 1946).

Propagated impulses can be set up in "slow" and "fast" (PLATE 3, FIGURE 3) muscles, whenever e.p.p.'s have grown or summed to a threshold level. Excitation in a "slow" muscle (opener of crayfish claw) is illustrated in PLATE 3, FIGURE 4. Stimulation at 57 p. sec. causes development of tension at a low rate, due to local shortening at the neuromuscular junction. The e.p.p.'s increase slowly to a steady small amplitude (PLATE 3, FIGURE 4a). When raising the frequency to 137 p. sec., the individual e.p.p.'s increase rapidly, and, at the same time, summation occurs. Although no propagated responses were set up, considerable muscle tension developed at a rapid rate (PLATE 3, FIGURE 4b). Further increasing the stimulating frequency to 160 p. sec. sets up a propagated muscle response.

2. Inhibition

The simplest preparation for the study of inhibition is the opener muscle of the crayfish claw, which is innervated by one motor and one inhibitory fiber. There are two different mechanisms by which muscle contractions can be inhibited:

a. *α-action (electrical inhibition):* The electrical response of the muscle is reduced to a variable extent, depending upon the relative time of arrival of the inhibitory and motor impulses at the neuromuscular junction. If the inhibitory impulse precedes the excitatory by several msec. at a steady stimulation of 50 per sec., the e.p.p.'s may be reduced as much as 70 to 80 % (PLATE 3, FIGURE 5 $A_{(4)}$, $B_{(2)}$). As the interval between the two nerve impulses is increased, the α-effect diminishes gradually (cf. Marmont and Wiersma, 1938). Since the local contractions at the neuromuscular junctions are determined by the e.p.p. size, it is clear that the α-action will reduce or prevent contraction.

b. *β-action (direct mechanical inhibition):* However, even if the inhibitory nerve impulse arrives at the neuromuscular junction too late to affect the preceding e.p.p., and too early to reduce the following one, it will still abolish contraction which normally follows motor nerve stimulation (PLATE 3, FIGURE 5 $B_{(3)}$). This important effect was discovered by Marmont and Wiersma (1938). It is, thus, possible to inhibit "excitation" of the contractile elements without altering the membrane potential. This applies only to *local* neuromuscular contractions. There is good reason to believe that inhibitory and motor nerves end in close proximity, and it seems that the inhibitory effect is confined to the neuromuscular region. There is no evidence that an already propagating muscle impulse and its associated contraction can be inhibited. (For details of α and β-action, see Kuffler and Katz, 1946).

DISCUSSION

Gradation of Muscular Contraction in Vertebrates and Crustaceans

Crustaceans, with one or a few motor nerves supplying whole muscles, can grade their muscular reactions in even finer steps than vertebrates with ample nerve supply. (i) The intensity and speed of local contractions in muscle fibers depend on the rate and the size of e.p.p.'s built up at the junctions. Not only does an increased number of stimuli increase the size of the individual potentials at a given region, but also a larger area becomes involved by electrotonic spread. Thus, facilitation and summation of local processes itself provide a wide range of activation at myoneural junctional regions. (ii) Apart from local gradation of contractions, *propagated* impulses can be set up at a variable rate in a variable number of muscle fibers. These, again, are

regulated by the number and frequency of nerve impulses (PLATE 3, FIGURE 4).

By a combination, therefore, of summation of local graded responses and by facilitation of propagated impulses, very fine adjustments of muscular activity can be effected. In addition, the crustaceans have at their disposal the inhibitory mechanism.

In vertebrate muscles, on the other hand, the muscular contraction is graded in relatively large steps, by the all-or-none responses of motor units. Although the contractile elements of voluntary vertebrate muscle are capable of graded contractions, as in crustaceans, normally they are maximally activated by the individual muscle impulses (Hill, 1938; Gerard, 1941; Brown, 1941; also discussed by Katz and Kuffler, 1946). Only when the threshold of the muscle membrane for propagated impulses is increased by drugs or by fatigue, can the action of local potentials become effective. The situation seems different when stimulation is done with microelectrodes. The earlier observations on graded responses, by Gelfan and Gerard (1930), and Gelfan and Bishop (1932), were probably due to excitation of a small membrane area not large enough to set up propagated impulses. In the absence of accurate recordings from the region of those contractions, possibly from the stimulating cathode (Kuffler, 1942, Figure 11), it may not be justified to conclude that no membrane changes occurred.

Whether produced by nerve impulses or by applied catelectrotonic potentials, responses of the muscle system seem to depend on the state of the membrane. The local contractions are the result of local potentials, both being capable of gradation; the propagated contractions set up by the muscle impulses, however, are the result of an intensive, probably maximal activation, which cannot be graded.

Action Currents and Contractile Elements

It seems to be widely assumed that the muscle membrane initiates the processes in the contractile elements by way of action currents which are produced during the impulse or during depolarization in general. From the study of contractures, it appears that action currents are not necessarily the essential link. When contractures are produced by application of drugs, the membrane resting potential is reduced, and the area becomes relatively negative to the rest of the fiber. The main direction of current flow in such a case will be from the normal positive area into the negative region of contracture. This current does not cause the contracture, since further addition to it in the same

direction, by making it the anode of a constant current pulse, will inhibit the contracture. Further, in a muscle which is entirely immersed into a contracture-producing solution, the whole membrane will be simultaneously depolarized. It is clear that, in such cases, no current flow from the "normal" to a depolarized region can occur. It seems that the essential condition for the initiation of processes in the contractile mechanism, is the "removal" of at least part of the membrane (within a certain time). This can be done at the cathode of a constant current, or by chemical means. Any process restoring the membrane, *e.g.*, an anodal pulse, will also cause relaxation, since it removes the condition which, in the first place, started the process. The depolarized state is apparently also necessary for the maintenance of the contracture. Therefore, "cathodal contractures" last for the approximate duration of the current flow, and chemical contractures (PLATE 2, FIGURES 1 and 2) last for the duration of the membrane changes. The question of what the link is between the membrane and the contractile elements, naturally arises. At present, only speculation seems possible. One may think, for instance, of a similar arrangement to that found in crystals, *i.e.*, the existence of an energy level system as suggested by Szent-Györgyi (1941). In crystals, removal of some electrons would throw the remainder of the system into activity. The question as applied to a biological system would be: How is the energy liberated by one molecule communicated to a greater number of similar molecules? Such a scheme again raises difficult questions which could not be explained without further knowledge of the molecular pattern in muscle.

A possible clue in the search for the membrane-contractile system link may come from the study of nervous inhibition of contraction. By interfering between the membrane and the contractile elements, the inhibitory impulse, if suitably timed, may prevent the normal excitatory effect of a potential change (PLATE 3, FIGURE $5B_{(3)}$). Here, for the first time, it was possible to separate the membrane changes from the contractile elements (Marmont and Wiersma, 1938). Such inhibition, however, applies only to local, non-propagated potentials around the neuromuscular junction. Propagated muscle impulses without contraction have not been observed (Kuffler and Katz, 1946).

This inhibitory action is, at present, quite mysterious. The inhibitory nerve impulse is conducted in a way similar to the motor impulse. Although it ends in close proximity to the endplate and, in addition, probably penetrates the membrane, as indicated by its effect, it does not cause a potential change.

SUMMARY

The role of the muscle membrane in bringing about and maintaining activity of the contractile elements is discussed.

Evidence is presented that the muscle shortening is set up by the membrane changes. Contractures, like propagated contractions, are initiated after a sufficient depolarization has been obtained. No essential difference is seen between the mechanism setting up both processes.

In vertebrate muscles which give contractures, and in crustacean muscles, a correlation is found between the extent of membrane potentials and the activity of the contractile system.

The inhibition of contractures by nerve impulses in crustaceans and by current pulses in vertebrate muscles, is discussed.

It is suggested that action currents are not the essential link which transmits "excitation" between the muscle membrane and the contractile elements.*

LITERATURE CITED

1. **Acheson, G. H., & A. Rosenblueth**
 1941. Some effects of veratrine upon circulated mammalian nerves. Am. J. Physiol. **133**: 736–751.
2. **Biedermann, W.**
 1896. Electrophysiology. Macmillan & Co. London.
3. **Bremer, F.**
 1932. Researches on the contracture of skeletal muscle. J. Physiol. **76**: 65–94.
4. **Brown, D. E. S.**
 1941. The regulation of energy exchange in contracting muscle. Biol. Symp. **3**: 161–190.
5. **Cowan, S. L.**
 1940. The action of eserine-like compounds upon frog's nerve-muscle preparations, etc. Proc. Roy. Soc. London B **129**: 356–411.
6. **Eccles, J. C., B. Katz, & S. W. Kuffler**
 1941. Nature of the "endplate potential" in curarized muscle. J. Neurophysiol. **5**: 362–387.
7. **Eccles, J. C., B. Katz, & S. W. Kuffler**
 1942. Effect of eserine on neuromuscular transmission. J. Neurophysiol. **5**: 211–230.
8. **Eccles, J. C., & W. J. O'Connor**
 1941. Abortive impulses at the neuromuscular junction. J. Physiol. **100**: 317–328.

* In recent investigations, a new system of neuromuscular excitation of skeletal muscle, differing from that described in this paper, has been studied. Small diameter nerve fibers, when stimulated repeatedly, set up local contractions around the junctional region. Electrically, the local changes resemble curarized endplate potentials or local potentials of normal crustacean muscles (Section B, above). The small-nerve motor system is active in reflexes and may produce appreciable muscle tension which can be graded in much finer steps than the propagated contractions of the well-known motor units. A distinct reflex pattern in the spinal cord has been found. (For references, see **Kuffler, S. W.**, Proc. Soc. Exp. Biol. & Med. **63**: 21–23. 1946; and **Kuffler, S. W., & R. W. Gerard**, J. Neurophysiol. In press. 1947.)

9. **Feng, T. P.**
1938. Further observations on propagation of veratrine contracture. Chin. J. Physiol. **13**: 239–246.

10. **Feng, T. P.**
1941. The local activity around the skeletal N.-M. junction produced by nerve impulses. Biol. Symp. **3**: 121–152.

11. **Gasser, H. S.**
1930. Contractures of skeletal muscle. Physiol. Rev. **10**: 35–120.

12. **Gelfan, S., & R. W. Gerard**
1930. Studies on single muscle fibres. II. A further analysis of the grading mechanism. Am. J. Physiol. **95**: 412–416.

13. **Gelfan, S. & G. H. Bishop**
1932. Action potentials from single muscle fibres. Am. J. Physiol. **101**: 678–685.

14. **Gelfan, S.**
1933. The submaximal responses of single muscle fibre. J. Physiol. **80**: 285–295.

15. **Gerard, R. W.**
1941. The Body Functions. John Wiley & Sons. New York.

16. **Graham, H. T., & H. S. Gasser**
1931. Modification of nerve response by veratrine, protoveratrine and aconitine. J. Pharm. & Exp. Therap. **43**: 163–185.

17. **van Harreveld, A., & C. A. G. Wiersma**
1939. The function of the quintuplet innervation of a crustacean muscle. J. Exp. Biol. **14**: 121–133.

18. **Katz, B.**
1939. Electric excitation of nerve. Oxford University Press. London.

19. **Katz, B., & S. W. Kuffler**
1946. Excitation of the nerve-muscle system in crusteacea. Proc. Roy. Soc. London B **133**: 374–389.

20. **Kuffler, S. W.**
1942a. Responses during refractory period at the myoneural junction in isolated nerve-muscle fibre preparation. J. Neurophysiol. **5**: 199–209.

21. **Kuffler, S. W.**
1942b. Further study on transmission in an isolated nerve muscle fibre preparation. J. Neurophysiol. **5**: 309–322.

22. **Kuffler, S. W.**
1943. Specific excitability of the endplate region in normal and denervated muscle. J. Neurophysiol. **6**: 99–110.

23. **Kuffler, S. W.**
1945. Action of veratrine on nerve-muscle preparations. J. Neurophysiol. **8**: 113–122.

24. **Kuffler, S. W.**
1946. The relation of electric potential changes to contracture in skeletal muscle. J. Neurophysiol. **9**: 367–377.

25. **Kuffler, S. W., & B. Katz**
1946. Inhibition at the nerve muscle junction in crustacea. J. Neurophysiol. **9**: 337–346.

26. **Marmont, G., & C. A. G. Wiersma**
1938. On the mechanism of inhibition and excitation of crayfish muscle. J. Physiol. **93**: 173–193.

27. **Szent-Györgyi, A.**
1941. The study of energy levels in biochemistry. Nature **148**: 157–159.

28. **Young, A. C.**
1941. Action potentials of skeletal muscle. Biol. Symp. **3**: 152–160.

PLATES 2–3

PLATE 2

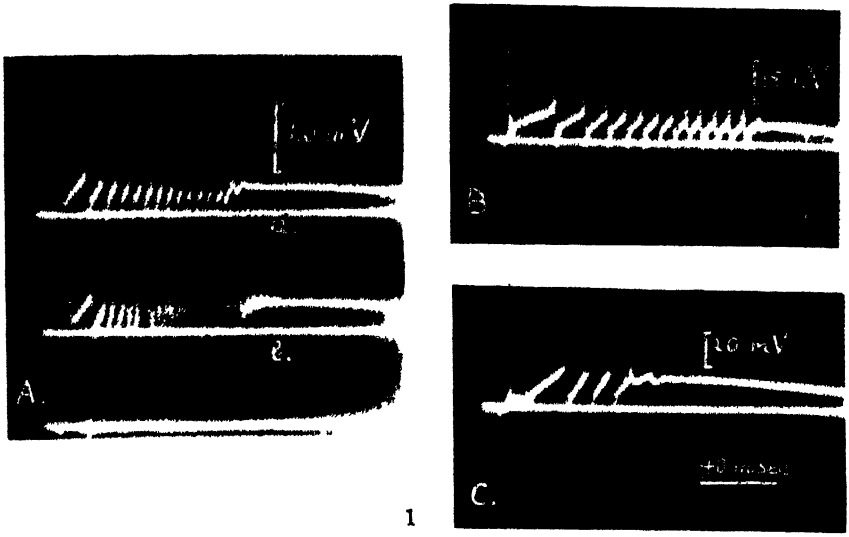
FIGURE 1. Electric potential changes recorded from three different preparations containing two muscle fibers. A small drop of veratrine (concentration 10^{-5} and 10^{-6}) was applied to the region in contact with one recording electrode. Then the muscle fibers were electrically stimulated, and the impulses passing over the veratrinized region initiated a series of discharges which were recorded at the site of their origin.

A. (a) and (b), recorded at an interval of half a minute. The two muscle fibers discharging almost simultaneously. One fiber ceases to discharge abruptly, while the other fiber continues to set up small responses which do not reach the second recording electrode. The following smooth negative potential may persist for several seconds.

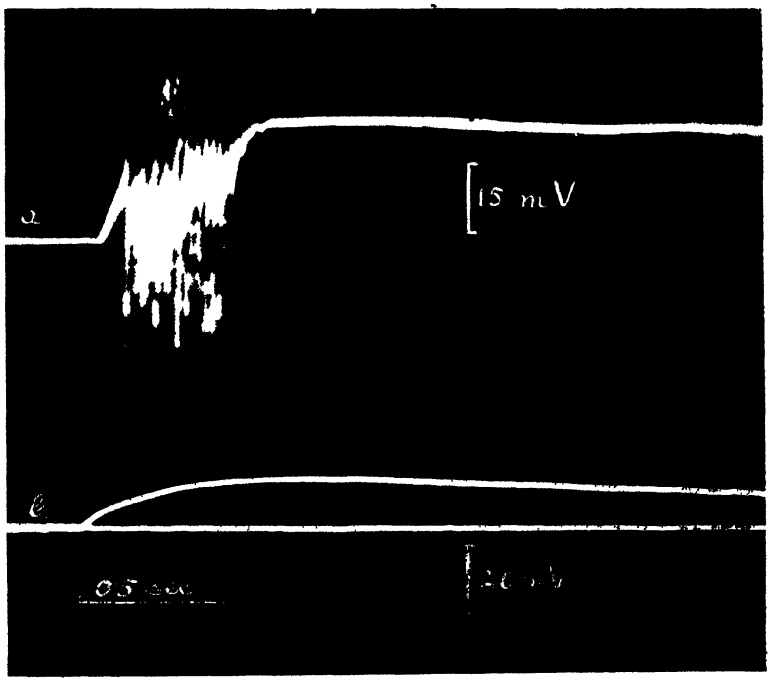
B. Note the gradual decrease of the muscle spike height in the veratrine region while the diphasic component remains constant.

C. The first two potentials are set up by stimulation of the muscle fibers at an interval of 8 msec. At the veratrine region, they initiate three fully propagated responses and some abortive impulses. (From Kuffler, 1946.)

FIGURE 2. (a), recording from the region of local application of 10^{-5} ACh to several muscle fibers. Note the small local responses following the propagated muscle impulses. The maintained negative potential is accompanied by contracture. (b), potential change in two muscle fibers set up by 0.6% KCl producing contracture. The apparent potential decay is due to the relatively short time constant of the amplifier. (From Kuffler, 1946.)

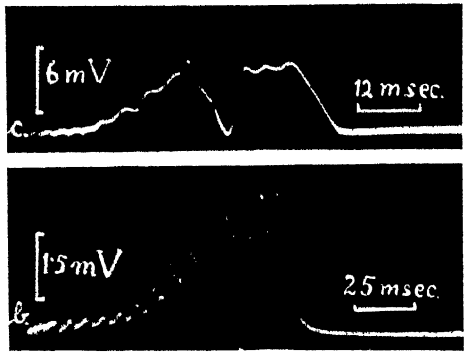


1



2

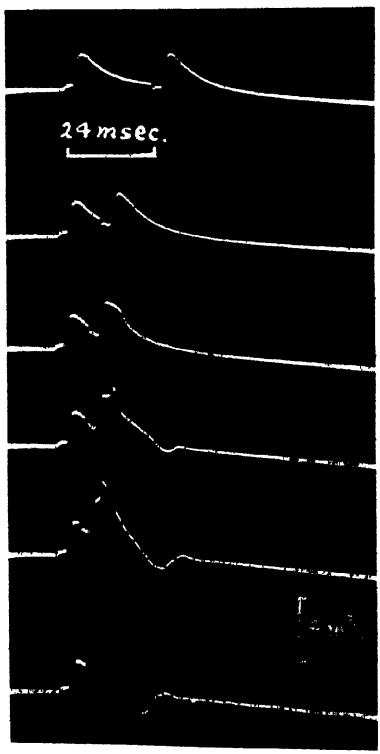
KUFFLER: MEMBRANE CHANGES



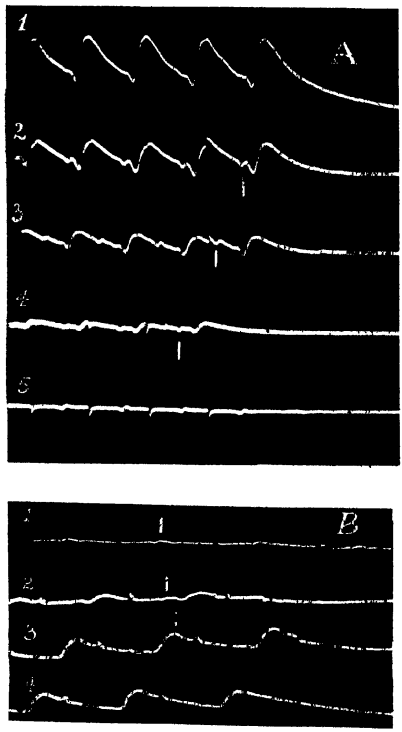
4



3



5



KUFFLER: MEMBRANE CHANGES

PLATE 3

FIGURE 3. "Fast" system. E.p.p.'s and propagated impulses recorded from junctional focus of extensor of the carpopodite. "Strip" preparation, innervation restricted to several muscle fibers. Both electrodes on strip. Nerve stimulation: intervals between shocks, successively from above, 24, 12, 9.6, 8, 6.4, and 4.8 msec. Note propagated spikes taking off at about 4 mV. Abortive spike at 9.6 msec. interval. (From Katz & Kuffler, 1946.)

FIGURE 4. "Slow" system. Opener of claw. Recording from a junctional focus. Nerve stimulation. Lower record, 57 per sec. (a), initial stimuli; (a, 1), after 0.2 sec. continual stimulation. Note intramuscular nerve spike of uniform size precedes e.p.p. (b), 135 per sec. (c), 160 per sec. Initial spike followed by local "plateau." (From Katz & Kuffler, 1946.)

FIGURE 5. Opener of claw, recording from junctional focus of exposed muscle. Stimulation at 50 per sec. A. (1) Motor nerve stimulation alone, accompanied by contraction. (2–4) Motor and inhibitory stimuli at different intervals (no contraction). Position of last inhibitory nerve spike is indicated by dashes. Maximum α-action at (4). (5) Inhibitory nerve stimulated alone. Note small spike indicating arrival of inhibitory nerve impulse at junctional region.

B. (1) Inhibitory stimulation only. (2) Inhibitory nerve impulse preceding motor impulse by several msec. (3) No appreciable α-action, mechanical inhibition complete. (4) Motor stimulation only (contraction). Inhibitory nerve spikes shown by dashes. From Katz & Kuffler, 1946.)

MUSCULAR CONTRACTION

PART II

ULTRASTRUCTURE

BIREFRINGENCE AND ULTRASTRUCTURE OF MUSCLE

By Ernst Fischer

Baruch Center of Physical Medicine, Medical College of Virginia, Richmond, Virginia

Birefringence, electron microscope, and x-ray diffraction studies offer three independent approaches to a visualization of the ultrastructure of the muscle. The structural basis of the muscle machine is of the utmost importance for any theory of the contractile mechanism. The dynamics of the contractile mechanism, in regard to elastic properties and tension production, have been discussed in the preceding papers (Ramsey, 1947; Gilson *et al.*, 1947). Proposed visualizations of the muscular ultrastructure derived from any of the three approaches mentioned above should not be contradictory to one another, and should satisfy the known dynamics of the muscle.

X-ray diffraction and electron microscopy are methods developed only recently, and the results derived from those methods on muscles will be discussed in the following paper by F. O. Schmitt and coworkers (1947). The physical phenomenon of double refraction or of birefringence, as it is more often called nowadays, has been well known since the classical investigation of Huygens (1690), in the last quarter of the 17th century. Nevertheless, it was only in 1858 that the fact that muscle is birefringent was discovered by E. Brücke. The latter explained this birefringence by the existence, within the muscle, of submicroscopical elongated particles arranged parallel to the long axis of the muscle, particles which have a higher refractive index than the surrounding phase. Under the influence of Nägeli's general theory of biological submicroscopical units, these hypothetical particles in the muscle have been called *micellae*. Birefringence of such an origin as assumed by Brücke is called today *form birefringence* or *birefringence of a composed body,* and the mathematical relation between sign and strength of birefringence and certain spatial and optical conditions of the micellar pattern has been developed by Wiener (1909, 1912) and tested experimentally for various non-biological material by Ambronn (1916, 1917, 1919). According to these studies, the birefringence, as observed on muscle, can originate from two different micellar patterns. In both, the long axis of the micellae are arranged parallel to the long axis of the muscle, but in one of the patterns the micellae are rodlets,

while in the other they are thin platelets. Incidental evidence favors strongly the assumption of rod-shaped micellae, but the cross-section of such an elongated micella is not necessarily an exact circle.

W. J. Schmidt (1924a, 1937) and others realized that many biological objects also exhibit birefringence due to a true crystalline structure, and that, in some objects with a micellar pattern, the *total birefringence* observed is due to *form birefringence*, as well as to *crystalline birefringence*, caused by the anisotropic molecular structure of the micellae themselves. As we will see later, muscle belongs to this class of biological objects.

Total birefringence in its sign and strength is easily determined, by compensation methods, as the phase difference between polarized light waves passing through planes parallel to the longitudinal axis of the muscle and the light waves passing through planes perpendicular to the former. Since this phase difference increases proportionally with the thickness of the material, birefringence is measured as phase difference per mm. thickness of the material under investigation. In calculating the birefringence of a muscle, the error in the determination of the thickness is often distinctly larger than the error in the determination of the observed phase difference.

Birefringence observations on skeletal muscle are complicated by the fact that its fibers are built up by sarcomeres consisting of two discs of different optical properties, thus causing the cross-striation. Contrary to the general statement in most text-books, the so-called isotropic disc is not truely isotropic but is slightly birefringent, while the anisotropic disc is highly birefringent, as shown by W. J. Schmidt (1934, 1935). To avoid the complications due to the cross-striation of skeletal muscle, many of the birefringence investigations were performed on smooth muscles.

From the measurements of total birefringence of living smooth and striated muscles (see TABLE 1), it becomes evident that the ultrastructure of smooth muscle is similar to, but not identical with, the ultrastructure of the anisotropic disc of skeletal muscle. The birefringence of striated muscle is only very little increased by elongation beyond that length at which all fibers are apparently straight and parallel. Buchthal and Knappeis (1938) have shown this for isolated fibers of the frog, and Fischer (1944) for the rectus abdominis of the mouse. However, elongation considerably increases birefringence in smooth muscles of invertebrates (Fischer, 1936a, 1938; Bozler and Cottrell, 1937) and of mammals (Fischer, 1944). Due to this distinct dependence of the birefringence of smooth muscle on the length, any exact

TABLE 1

TOTAL BIREFRINGENCE IN 10^{-3} OF SURVIVING STRIATED AND SMOOTH MUSCLES

Species	Striated muscle		Smooth muscle	
dog[1]	sartorius	2.60	intestinal muscle	2.92
	rectus abdominis	2.57	retractor penis	3.33
mouse[1]	rectus abdominis	2.57		
frog	sartorius[2]	1.70		
	sartorius[1]	2.24		
	sartorius[3]	2.50		
snail[4]			retractor	2.00
Phascolosoma			retractor[4]	1.90
			retractor[5]	2.65
Thyone[5]			retractor	1.55

[1] **Fischer**, 1944; [2] **Buchthal & Knappeis**, 1938; [3] **Ebner**, 1882; [4] **Bozler & Cottrell**, 1937; [5] **Fischer**, 1938.

comparison between birefringence of smooth and striated muscle is difficult (see FIGURE 1).

The relation between muscle length and birefringence is complicated, in smooth muscles, by their tonic and plastic properties, in consequence of which, no fixed relation between muscle length and tension exists. That birefringence is apparently dependent on length and on tension for invertebrate smooth muscles, can be clearly demonstrated in experiments with alternating stretching and releasing of the muscle with increasing loads (Fischer, 1938). Retractor muscles of *Phascolosoma,* a marine worm, with the cerebral ganglion still attached, or retractor muscles of *Thyone,* a holothurian, are especially suitable for such experiments, since these muscles relax their tonus rather slowly after preparation. Therefore, it is possible also to measure birefringence for relatively short muscle length. If invertebrate smooth muscle preparations with no or little tone are stretched at a slow rate, birefringence increases with length exactly as birefringence increased for the released condition in the experiment of FIGURE 1, A and B (Bozler and Cottrell, 1937; Fischer, 1938). Bozler pointed out that birefringence of the relaxed retractor of the snail increases linearly with the square root of the muscle length. This relation seems to be true for all invertebrate smooth muscles, as far as yet tested, and, as Bozler emphasized, this relation could be accounted for, if the elements causing birefringence were arranged in surfaces. However, this relation between birefringence and muscle length does not hold true for mammalian smooth muscle, such as the retractor penis of dogs, for which the birefringence increase with stretch is considerably less than that demanded by the square root rule. As mentioned before, the increase

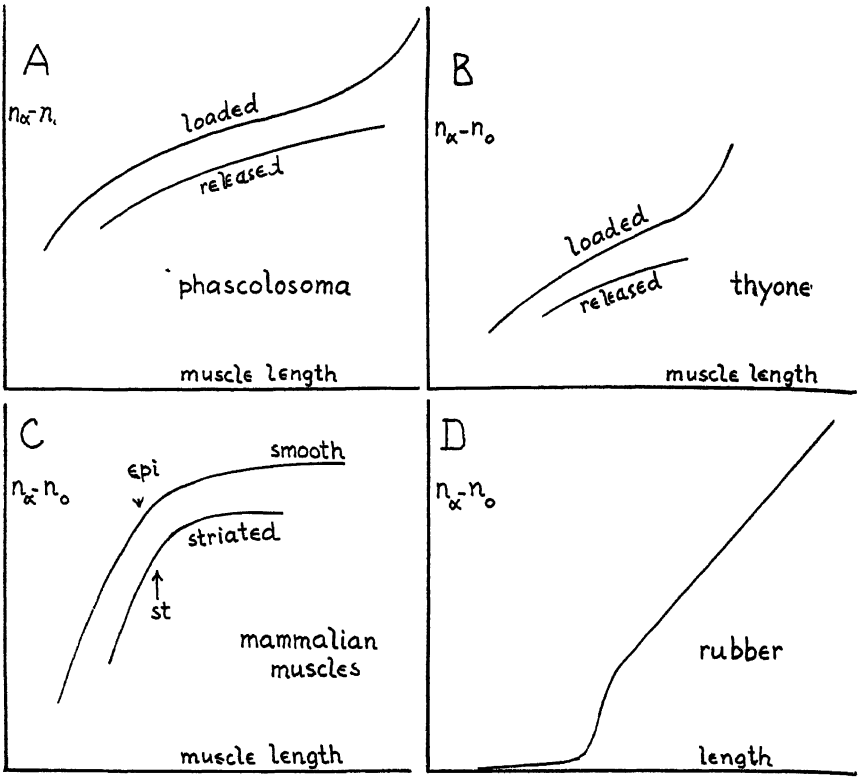


FIGURE 1. Relation between length and birefringence: A for retractor of *Phascolosoma* (a marine worm), under tension (loaded), and under no tension (released); B for retractor of *Thyone* (a holothurian), under tension (loaded), and under no tension (released); C for the smooth retractor penis of the dog and for the striated rectus abdominis of the mouse (values left of the arrows measured after activation by electrical stimulation, st, or by epinephrine, epi; D for crepe rubber (according to **Treloar,** 1941).

in birefringence with length for striated mammalian muscle is still smaller than for the mammalian smooth muscle.

All recent investigators agree that, during isotonic contraction, total birefringence of both smooth and striated muscles diminishes considerably during shortening and returns to initial values during relaxation. Under isometric conditions, appreciable birefringence decreases occur for striated muscle at low initial length. The time course of the birefringence change corresponds exactly to the mechanogram (Bozler and Cottrell, 1937). This decrease in birefringence during contraction becomes smaller and smaller with increased initial length, so that, finally, for a very much stretched muscle, the birefringence decrease is very small and difficult to measure. According to records published

by von Muralt (1932), the isometric birefringence decrease at low initial length is a three-phasic variation: a quick decrease during the beginning of tension development, a partial return and maintenance of an only slightly decreased birefringence, and a second quick decrease during relaxation, with return to the initial birefringence about the time relaxation is completed. However, his observations could not be confirmed, and were probably due to the use of the pelvic end of the sartorius, since the sarcomeres of the pelvic end show, during a single isometric twitch at low initial tension, alternatingly, shortening by their own contraction and elongation by stretch from other stronger parts of the same muscle (Fischer, 1926; Sandow, 1936). Buchthal (1940) reported for single cross-striated fibers that the decrease in birefringence during isometric contraction shows even three maxima of equal heights at about 15, 40, and 60 msec. after single stimulation. However, no description of the method employed and of the mechanical conditions has yet been furnished by Buchthal, and it is impossible, therefore, to evaluate this report.

For smooth muscles, probably due to their larger extensibility, it is much easier to demonstrate that, with increasing resting length, the birefringence decrease during isometric contraction diminishes. It is well established that smooth muscles at high initial length show no decrease of birefringence during isometric contraction and even that, occasionally, small increases are observed (Fischer, 1936a, 1944; Bozler and Cottrell, 1937). It must be pointed out here that these increases are so small that they could be simulated by the errors in the determination of the thickness of the resting and stimulated muscles. However, statistical evaluation, from my own material, of all data for isometric contraction at maximal length reveals, on the average, an increase of $2.3 \pm .5\%$, which is statistically significant.

All these measurements of total birefringence on living muscles are more or less in agreement with the results of x-ray diffraction studies, if we assume that the observed birefringence is crystalline birefringence as defined earlier. The main factor governing the degree of orientation of the molecular space-lattice is muscle length. In striated muscle, maximal orientation is nearly reached at normal length, and further stretch perfects orientation of the molecular chains very little, while elongation of the muscle is mainly due to slipping of structural units. Tension development affects birefringence mainly by its influence upon muscle length. Only in invertebrate smooth muscle where, apparently, no sliding of chain molecules can occur at maximal length, can a weak photoelastic effect be produced by tension. The small influence of

stretch upon the birefringence of muscle is in contrast to the distinct increase in birefringence of rubber. According to Treloar (1941), this birefringence increase of rubber, which is beyond a minimal length proportional to elongation (FIGURE 1D), is caused mainly by increased crystallization due to the elongation.

Schmidt (1924b) developed an imbibition method which enables one to find out, even for biological objects, how much of the total birefringence is due to form birefringence and how much is due to crystalline birefringence. This method can be used only on fixed muscles. With a careful choice of the fixation fluid, the difference between the total birefringence, measured before and after fixation of smooth and striated muscle, is negligible (Noll and Weber, 1935; Fischer, 1936a, 1938, 1944). There is no doubt that the muscle fibers shrink somewhat with even the best fixation method, and the apparently unchanged birefringence is due to a simultaneous decrease in anisotropic orientation which just compensates for the increase in birefringence due to shrinkage. The principle of the imbibition method consists of replacing the natural surrounding phase of the micellae by fluids of various refractive indices. Form birefringence is abolished as soon as the micellae are surrounded by a phase with the same refractive index as that of the micellae themselves. Form birefringence becomes apparent again for fluids with higher refractive indices than the micellae. If the birefringence for any material does not become zero for any imbibition fluid, but assumes only a minimum with a certain fluid, this rest birefringence must be caused by the crystalline birefringence of the micellae themselves.

That the application of such an imbibition method to fixed muscles gives reproducible results, is rather astonishing. Taking these results at their face value (TABLE 2), they would indicate that our previous assumption that total birefringence is caused only by the molecular space-lattice is incorrect. The results indicate clearly a micellar pattern for the muscles. However, the reproducibility of the results of the imbibition investigation on muscle should not be taken as an assurance that their interpretations are as unambiguous as the interpretation of imbibition experiments with inorganic material (Ambronn, 1916, 1917, 1919). Whatever the correct interpretation might be, the results indicate clearly that the various muscles, as well as the artificial myosin thread, must have an ultrastructure which, in its main features, is the same for all of them and varies only in minor details. If we use the interpretation of Ambronn, the ratio form to crystalline birefringence is somewhat higher for striated muscle than for smooth

TABLE 2

BIREFRINGENCE VALUES IN 10^{-3}

Material	Total bire-fringence	Form bire-fringence	Crystalline bire-fringence	Refractive index of micellae
Myosin thread[1]	3.5	1.9	1.6	1.56
Sartorius, frog[2]	2.4	1.6	.8	1.57
Gastrocnemius, rat[3]	2.4	1.5	.9	1.57
Rectus abdominis, mouse[3]	2.6	1.8	.8	1.57
Sterno-hyodeus, dog[3]	2.5	1.7	.8	1.57
Retractor penis, dog[3]	3.2	2.3	.9	1.55
Intestinal muscle, dog[3]	2.7	1.8	.9	1.55
Retractor, *Phascolosoma*[4]	2.6	1.4	1.2	1.52
Retractor, *Thyone*[5]	1.9	1.2	.7	1.52

[1] **Weber, 1935**; [2] **Noll & Weber, 1935**; [3] **Fischer, 1944**; [4] **Fischer, 1936a**; [5] **Fischer, 1938.**

muscle in general. The refractive index of the micellae is lower for smooth than for striated vertebrate muscles. The lower value of 1.55 for the smooth vertebrate muscles is, however, still higher than that of 1.52 found for invertebrate smooth muscles. These differences might indicate either differences in the myosin of the various muscles and/or differences in the molecular arrangement of the myosin chains inside the micellae. The relatively high value of 1.2 for the crystalline birefringence of the *Phascolosoma* retractor might be due to the fact that this muscle is the only one of TABLE 2 which contains some paramyosin, according to Schmitt *et al.* (1947). It is possible to calculate roughly, from the absolute value of form birefringence and from micellar refractive index, the theoretical volume percentage of the micellar phase (Wiener, 1909; Bear *et al.*, 1937). This value is 7 per cent for the retractor penis, 5 per cent for the gut muscles of the dog and for the smooth *Phascolosoma* retractor, and 3 per cent for mammalian and frog striated muscles, as well as for the smooth *Thyone* retractor. From Noll and Weber's investigation (1935) it is obvious, however, that these figures represent only minimum values for the micellar volume, since, in the case of muscle, the penetration of the imbibition fluid is apparently not completely restricted to the intermicellar spaces, but advances somewhat into the micellae themselves.

With this imbibition method, also, the influence of stretch and shortening on the two components of total birefringence can be determined on suitable muscles (Fischer, 1936a, 1938, 1944). However, in evaluating such experiments, one must keep in mind that the muscles are not fixed in a true resting state nor in true contraction, but in a state of weak contracture. For invertebrate smooth muscle, elongation be-

yond resting length increases total birefringence, mainly due to an increase of form birefringence. From this, the conclusion can be drawn that, in the muscle at normal length, the micellar arrangement is not optimal for form birefringence, but improves with elongation, while the molecular orientation inside the micellae is only little improved by stretch. The decrease in total birefringence observed in the shortened state is due to almost complete disappearance of form birefringence, while crystalline birefringence is more than halved. From the results of such measurements on a series of comparable muscles, birefringence-length diagrams can be constructed (FIGURE 2). For all smooth muscle, as far as yet investigated, crystalline birefringence of the micellae increases slightly, more or less proportionally with length, when elongated beyond a certain minimal length, while

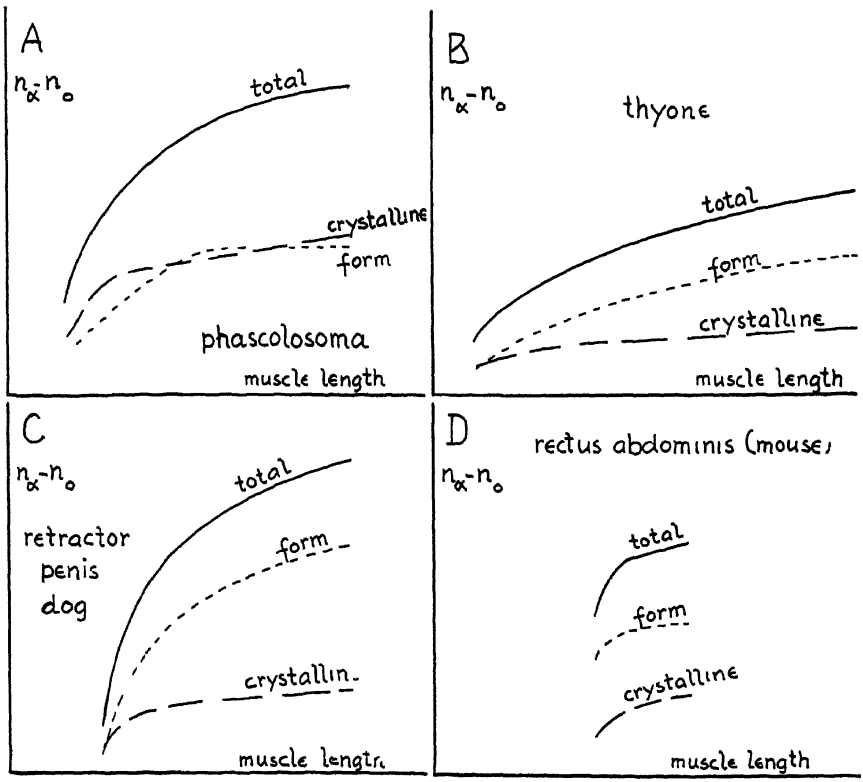


FIGURE 2 Total form and crystalline birefringence in relation to muscle length. A for retractor of *Phascolosoma;* B for retractor of *Thyone;* C for smooth retractor penis of the dog; D for striated rectus abdominis of the mouse

form birefringence increases at a higher but steadily declining rate. Elongation of striated muscles beyond normal length increases crystalline birefringence more distinctly than form birefringence This difference between smooth and striated muscle might be interpreted as due to the larger plasticity of the smooth muscles. In the latter, the micellae are more movable than in the striated muscles, and elongation of the whole muscle improves, therefore, the molecular orientation of the micellae much less in smooth than in striated muscles.

Smooth and striated muscles fixed at a length shorter than a minimal critical length show a great loss in birefringence, which is caused by a strong diminution in crystalline as well as in form birefringence. This is best interpreted as a shortening of the micellae to a rather spherical shape, without complete disorientation of the intra-micellar molecular space-lattice. The data of such birefringence observations, in conjunction with data from other investigations furnishing estimates for probable dimensions of the micellae and for the intra-micellar distances, have been utilized to construct hypothetical schemes of the micellar pattern for various muscles at different conditions (Hürthle, 1931; Fischer, 1936b).

Such schemes have rather limited values, not only since some of their details are based on rather rough estimates of dimensions, but mainly because the whole assumption that the muscle represents a true "composed body" is questionable. Jordan (1935) emphasizes rightly, in his discussion of the viscous and plastic properties of invertebrate smooth muscles, the structural and viscous character of the intermicellar medium. The x-ray diffraction data and the electron microscope observations indicate, also, that the ultrastructure of the muscle is much more complex than a simple micellar pattern, according to which bundles of myosin molecule chains forming micellae are suspended in a surrounding phase free of any structure. The possibility that the rodlets are interconnected by anastomosing chains has been emphasized by Schmitt (1941).

The electron microscope photographs of single striated muscle fibrils, published recently by Hall *et al.* (1946), demonstrate clearly that the myofibrils are formed by bundles of myosin filaments extending continuously and in relatively straight lines through both the isotropic and the anisotropic disc. The filaments, despite a few interconnecting anastomoses, separate laterally and adhere to one another only at the region of the *Z* and of the *M* membranes. In the birefringence studies with the imbibition methods, the fluids are apparently able to penetrate

between these myosin filaments, thus permitting one to measure the birefringence of the filaments themselves as "crystalline" birefringence. The width of these filaments ranges from 50 to 250 Å, which is the same order of magnitude as assumed in the various hypothetical micellar schemes. Although the myosin filaments have an indefinite length in the myofibrils, there is no doubt that myosin molecules bundled together as micellae can be easily extracted from muscle (Weber, 1939; Hall *et al.*, 1946). The x-ray diffraction pattern of such extracted myosin spun into threads is simpler than that of fresh muscle (Astbury and Dickinson, 1940). However, the fact that myosin can be extracted in micellar form indicates, at least, a certain predisposition of the ultrastructure of the muscle to break down in such a way that circumscribed bundles of myosin molecules are readily released. Astbury and Dickinson (1940), pointed out that, according to their x-ray studies, a concept essential for the proper understanding of the elastic properties of biological fibers and similar long chain synthetic products, is the alternation of better organized parts and less organized parts caused by gradation in both size and perfection of chain bundles or crystallites. The true fiber axis period of myosin of at least 350 to 450 Å, as revealed by small-angle x-ray diffraction studies (Bear, 1945; Schmitt *et al.*, 1947), might indicate such alternation of better and less organized parts. I am inclined to assume that the micellae, as visualized by the birefringence studies, correspond to the more perfectly organized parts of the x-ray studies. According to Guth (1947), the ultrastructure of bulk rubber can be visualized as long chains interlinked by a few cross-bonds to a three-dimensional network in which finite clusters of more loosely connected chains are embedded. We might assume that such a picture, in reverse, might give us a good visualization of the ultrastructure of the filaments: an infinite and rather loosely organized network of chain molecules, in which definite clusters of chain molecules linked by a relatively large number of cross-bonds are embedded.

That the chain molecules of the network might not be identical with the molecules linked into clusters is indicated, at least for striated muscle, by the discovery of Straub (1942, 1943) that, by prolonged treatment with KCl solutions, myosin, not alone, but in combination with another protein (actin), can be extracted from muscle mash. The methods needed for isolation of pure actin from muscle suggest that this protein forms in the muscle fiber a framework to which the myosin molecule adheres. Szent-Györgyi (1943a, 1943b, 1946) regards this combination of myosin and actin, which he calls *acto-myosin*, as the

substrate of the contractile mechanism. However, in his scheme, the myosin micellae represent the prime mover of the system.

In the light of all these findings, it might be permissible to regard the differences in the hypothetical micellar pattern, as revealed by birefringence analysis for different muscle types at various conditions, as an indication of some corresponding differences or changes in the true ultrastructure of the various muscles.

Despite the incomplete information about the true ultrastructure of the muscle which can be gained by birefringence measurements, such studies can reveal certain relations between perfection of ultrastructure and functional state of the muscle. For normal rat gastrocnemii, there exists, as shown by Fischer (1940), a statistical relation between total birefringence and contractile power per weight unit, which can be expressed by the equation: birefringence $= k \times \log$ contractile power. In various types of muscular atrophy, the loss in contractile power depends on the loss in birefringence. In atrophies without birefringence diminution, no loss in power per weight unit occurs. The marked loss in total birefringence, which corresponds to the loss in muscle power per weight unit during denervation atrophy, is caused mainly, but not exclusively, by a diminished form birefringence. Electrical daily exercise of denervated muscles can considerably retard their weight loss, but has only a slightly retarding effect upon the muscle power per weight unit and, correspondingly, upon the birefringence diminution.

As mentioned before, in denervation atrophy, the power loss corresponds to the birefringence diminution. Since the power per weight unit declines only about 10 days after denervation of a rat muscle, appreciable loss in birefringence is also only observed so late. In recent experiments, we were able to demonstrate that the time-course of the loss in extractable myosin of a denervated rabbit muscle (Fischer and Ramsey, 1946a) is very similar to the time-course of the birefringence diminution (FIGURE 3). Both loss in extractable myosin and loss in birefringence are quite independent of the weight loss of the atrophying muscles. As mentioned before, electrical treatment of atrophying muscles decreases loss in birefringence and in muscle power to the same extent. We found now that this treatment will also influence the loss in extractable myosin in exactly the same manner. Electrical training of normal muscles, despite the hypertrophy produced, does not increase muscle power per weight unit or birefringence (Fischer, 1940), or the amount of extractable myosin (Fischer and Ramsey, 1946b).

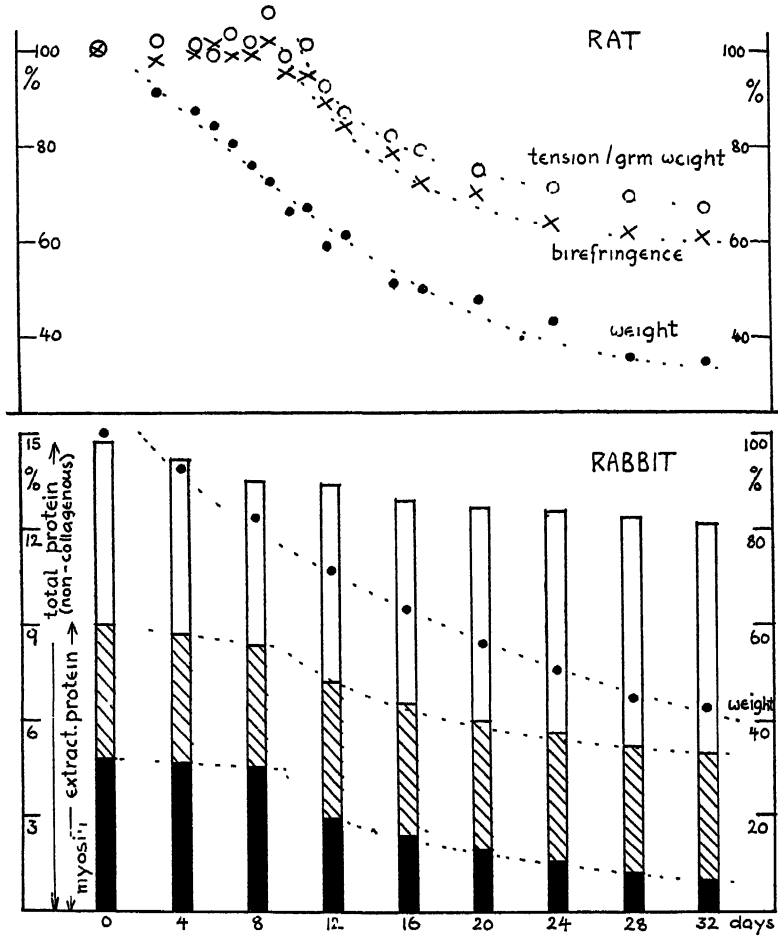


FIGURE 3. Upper half: weight, birefringence, and tension per gm. weight of denervated rat gastrocnemii. Lower half: weight, total non-collagenous protein, extractable protein, and myosin (all proteins expressed as per cent of muscle wet weight) of denervated rabbit gastrocnemii.

As can be realized from the material presented here, birefringence studies can contribute directly only little to our knowledge of the ultrastructure of the muscle, if compared with x-ray diffraction and electron microscope studies. However, combining the results of all these studies should lead to a visualization of the ultrastructure of the muscle compatible with all the data supplied by these methods, if we keep in mind that all these three approaches have their limitations and none of the interpretations is unambiguous.

The measurement and analysis of birefringence data offer at least

a relatively simple method for the detection of coarse changes in the ultrastructure of muscle, and the birefringence approach is more suitable than other methods to correlate functional changes with ultrastructure changes.

LITERATURE CITED

1. **Ambronn, H.**
 1916. Ueber das Zusammenwirken von Stäbchendoppelbrechung und Eigendoppelbrechung. Kolloid Ztschr. **18**: 90, 273.
 1917. Ueber das Zusammenwirken von Stäbchendoppelbrechung und Eigendoppelbrechung. Kolloid Ztschr. **20**: 173.
 1919. Ueber die akzidentelle Doppelbrechung im Zelloidin und in der Zellulose. Nachr. Kgl. Ges. Wiss. Göttingen, Math. phys. Kl.: 299.
2. **Astbury, W. T., & S. Dickinson**
 1940. X-ray studies of the molecular structure of myosin. Proc. Roy. Soc. London B **129**: 307.
3. **Bear, R. S.**
 1945. Small-angle X-ray diffraction studies on muscle. J. Am. Chem. Soc. **67**: 1625.
4. **Bear, R. S., F. O. Schmitt, & J. Z. Young**
 1937. The ultrastructure of nerve axoplasm. Proc. Roy. Soc. London B **123**: 505.
5. **Brücke, E.**
 1858. Untersuchungen über den Bau der Muskelfasern mit Hülfe des polarisierten Lichtes. Denkschr. kaiserl. Akad. Wiss. Math.-Naturw. Kl. **15**: 69.
6. **Bozler, E., & C. L. Cottrell**
 1937. The birefringence of muscle and its variation during contraction. J. Cell. & Comp. Physiol. **10**: 165.
7. **Buchthal, F.**
 1940. Experiments on the minute structure and function of the striated muscle fibre. Acta Physiol. Neurol. **15**: 43.
8. **Buchthal, F., & G. G. Knappeis**
 1938. Untersuchungen über die Doppelbrechung der einzelnen lebenden, quergestreiften Muskelfaser. I. Skand. Arch. Physiol. **78**: 97.
9. **von Ebner, V.**
 1882. Untersuchungen über die Anisotropie organischer Substanzen. W. Engelmann. Leipzig.
10. **Fischer, E.**
 1926. Die isometrische Muskelaktion des curarisierten und nichtcurarisierten Sartorius, seine Dehnbarkeit und die Fortpflanzung der Dehnungswelle. Pflüg. Arch. **213**: 352.
 1936a. The changes in birefringence during contraction and stretch of a smooth muscle (retractor of Phascolosoma). J. Cell. & Comp. Physiol. **8**: 503.
 1936b. The submicroscopical structure of muscle and its changes during contraction and stretch. Cold Spring Harbor Symp. Quant. Biol. **4**: 214.
 1938. The birefringence of smooth muscle (Phascolosoma and Thyone) as related to muscle length, tension, and tone. J. Cell. & Comp. Physiol. **12**: 85.
 1940. Changes during muscle contraction as related to the crystalline pattern. Biol. Symp. **3**: 211.
 1944. The birefringence of striated and smooth mammalian muscles. J. Cell. & Comp. Physiol. **23**: 113.
11. **Fischer, E., & V. W. Ramsey**
 1946a. Changes in protein content and in some physicochemical properties of the protein during muscular atrophies of various types. Am. J. Physiol. **145**: 571.

1946b. The effect of daily electrical stimulation of normal and denervated muscles upon their protein content, and upon some of the physicochemical properties of the protein. Am. J. Physiol. **145**: 583.

12. Gilson, A. S., Jr., G. M. Schoepfle, & S. M. Walker
1947. The time course of tension development in the muscle response. Ann. N. Y. Acad. Sci. **47**(6): 697.

13. Guth, E.
1944. Fundamental aspects of the elasticity of rubber and rubber-like material. Colloid Chemistry **5**: 287.
1947. Muscular contraction and rubber-like elasticity. Ann. N. Y. Acad. Sci. **47**(6): 716.

14. Hall, C. E., M. A. Jakus, & F. O. Schmitt
1946. An investigation of cross striation and myosin filaments in muscle. Biol. Bull. **90**: 32.

15. Hürthle, K.
1931. Zur Kenntnis der Struktur des ruhenden und des tätigen Froschmuskels. III. Ueber die Verteilung von Wasser und fester Substanz und über den submikroskopischen Bau der Fibrillen. Pflüg. Arch. **227**: 610.

16. Huygens, Christians
1690. Traité de la Lumière. Leyden.

17. Jordan, H. J.
1935. First Report on Viscosity and Plasticity. 214. Kon. Akad. Wetensch. Amsterdam.

18. von Muralt, A.
1932. Ueber das Verhalten der Doppelbrechung des quergestreiften Muskels während der Kontraktion. Pflüg. Arch. **230**: 299.

19. Noll, D., & H. H. Weber
1935. Polarisationsoptik und molekularer Feinbau der Q-Abschnitte des Froschmuskels. Pflüg. Arch. **235**: 234.

20. Ramsey, R. W.
1947. Dynamics of single muscle fibers. Ann. N. Y. Acad. Sci. **47**(6): 675.

21. Sandow, A.
1936. Diffraction patterns of the frog sartorius and sarcomere behavior under stretch. J. Cell. & Comp. Physiol. **9**: 57.

22. Schmidt, W. J.
1924a. Die Bausteine des Tierkörpers im polarisierten Licht. Fr. Cohen. Bonn.
1924. Anleitung zu polarisationsmikroskopischen Unterschungen für Biologen. Fr. Cohen. Bonn.
1934. Ueber die Doppelbrechung der I-Glieder der quergestreiften Myofibrillen und das Wesen der Querstreifung überhaupt. Ztschr. Zellforschg. **21**: 224.
1935. Nochmals über die Doppelbrechung der I-Glieder der quergestreiften Myofibrillen. Ztschr. Zellforschg. **23**: 201.
1937. Die Doppelbrechung von Karyoplasma, Zytoplasma und Metaplasma. Borntträger. Berlin.

23. Schmitt, F. O.
1941. Discussion of paper by **E. Fischer.** Biol. Sympos. **3**: 236.

24. Schmitt, F. O., R. S. Bear, C. E. Hall, & M. A. Jakus
1947. Electron microscope and x-ray diffraction studies of muscle structure. Ann. N. Y. Acad. Sci. **47**(6): 799.

25. Straub, F. B.
1942. Actin I. Studies Inst. Med. Chem. Univ. Szeged (Hungary) **2**: 3.
1943. Actin II. Studies Inst. Med. Chem. Univ. Szeged (Hungary) **3**: 23.

26. Szent-Györgyi, A.
1943. Observation on acto-myosin: 86; Note on acto-myosin: 98. Stud. Inst. Med. Chem. Szeged. **3**.
1946. Contraction and the chemical structure of the muscle fibril. J. Colloid Science **1**: 1.

27. Treloar, L. R.
1941. Crystallization phenomena in raw rubber. Trans. Faraday Soc. **37**: 84.

28. Weber, H. H.
1935. Der Feinbau und die mechanischen Eigenschaften des Myosinfadens. Pflüg. Arch. **235**: 205.
1939. Muscle proteins and properties of the muscle. Proc. Roy. Soc. London B **127**: 27.

29. Wiener, O.
1909. Zur Theorie der Stäbchendoppelbrechung. Ber. Verhdlg. Kgl. Sächs. Ges. Wiss. Math.-physik. Kl. **32**: 379.
1912. Die Theorie des Mischkörpers. Abhandlg. Sächs. Ges. Wiss. Math.-physik. Kl. **32**: 379.

ELECTRON MICROSCOPE AND X-RAY DIFFRACTION STUDIES OF MUSCLE STRUCTURE

By F. O. Schmitt, R. S. Bear, C. E. Hall, and M. A. Jakus

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts

In the study of the mechanism of muscle contraction, skeletal muscle has been the material most widely investigated. There are many physiological and chemical reasons for this, but the structural aspects are also important. The complex cytoarchitecture of cross-striated muscle and its alterations in various physiological states furnish visible clues regarding the contractile mechanism.

The anisotropy of the alternate bands early led to a recognition of the importance of molecular orientation in contraction (Engelmann, 1875). The demonstration that myosin solutions show double refraction of flow (Muralt and Edsall, 1930) led to a localization of this protein in the anisotropic (*A*) bands. However, subsequent, more or less indirect lines of evidence indicate that the entire myofibril is composed of myosin, while the relative isotropy of the *I* bands was ascribed to a lack of preferred orientation of myosin in these regions (Schmidt, 1937).

With the discovery of the enzymic relationship between myosin, adenosinetriphosphate (ATP), and the mineral constituents, attention has again been directed to the localization of these substances in the muscle fibril. Explanation of contraction and relaxation is currently sought in topochemical reactions between the structural protein, myosin, and ATP activated by mineral constituents.

ELECTRON MICROSCOPE OBSERVATIONS

Because of its high resolving power, the electron microscope offers unique opportunities as a means of investigating muscle structure. Ardenne and Weber (1941) have published electron micrographs of myosin particles extracted from skeletal muscle. Richards, Anderson, and Hance (1942) and Sjöstrand (1944) have employed striated muscle in an investigation of methods of obtaining electron micrographs of thin sections of tissue. However, the first detailed electron microscope study of the structure of myofibrils of striated muscle is that of Hall, Jakus, and Schmitt (1946). These investigations represent only

a beginning. Obviously, the detailed description of the structure of a tissue as complex as skeletal muscle will occupy many workers for years to come. Particularly significant will be the application to the problem of new techniques. However, the results already obtained have an important bearing on the theory of contraction and the properties of myosin. These aspects will be briefly considered below. For details of methods, the original paper may be consulted.

Myofibrils of Cross-Striated Muscle

The structures observed in electron micrographs of vertebrate and invertebrate myofibrils are readily identifiable with those described in the histological literature (Jordan, 1933), and correspond with them remarkably well. The myofibril is composed chiefly of fine filaments, 50 to 250 Å in width, extending in relatively straight lines parallel with the fibril axis (PLATE 4a). The *Z* membranes, which demarcate the sarcomeres, appear in the form of dense, rather amorphous material extending across the fibril. The *A* bands are differentiated from the *I* bands by the presence in them of relatively dense material which combines with phosphotungstic acid which was used as an electron stain. This material may be chiefly mineral in nature. There is evidence that calcium, magnesium, and potassium are preferentially located in the *A* bands (MacCallum, 1908; Winter and Smith, 1922; Scott, 1932; Scott, 1939; Dubuisson, 1942). However, for the present, the material is referred to as "*A* substance." The boundary between the *A* and *I* bands is sharp, though no membranous or other structural enclosure has been discerned. The *M* and *N* bands, as well as the *H* discs, are also apparent.

Collagen fibrils are associated with the myofibrils, but the association is limited to the periphery of the fibrils and the interfibrillary space. The fibrous matrix of the myofibrils is composed of the longitudinal filaments extending without apparent interruption through the consecutive *A* and *I* bands. In view of the high concentration of myosin in muscle fibrils (Smith, 1937), it is reasonable to suppose that the filaments are composed principally of myosin. They will, therefore, be referred to as *myosin filaments*.

The above description refers to myofibrils obtained from muscles kept extended during fixation. Observations were made also on fibrils from muscles which were slightly contracted or held in vigorous contraction by electrical stimulation during fixation (PLATE 4b). Isotonic as well as isometric contractions were thus induced. Characteristic "reversal of striation" (Jordan, 1933) was observed, *i.e.*, *A* sub-

stance moved toward the *Z* bands, away from the *M* bands. Contraction, thus, is accompanied by redistribution of *A* substance along the myosin filaments.

Particularly significant is the structure of the myosin filaments. In uncontracted fibrils, these filaments course through both *A* and *I* bands without apparent interruption, at least for the length of several sarcomeres, as observed in the limited field portrayed in the electron micrographs. The filaments are somewhat less perfectly aligned in the *I* bands than in the *A* bands. However, the relative isotropy of the fibrils in the region of the *I* bands cannot be explained on the basis of random orientation of the myosin filaments in these regions.

In strongly contracted fibrils (50%), as indicated by the sarcomere length, the myosin filaments are as straight as in uncontracted fibrils. This can only mean that the filaments themselves shorten in contraction, rather than coil up or fold on themselves. The myosin filament is, therefore, the smallest contractile unit thus far observed with the electron microscope.

The myosin filaments manifest discontinuities along their length, giving a nodose appearance. This is particularly evident in stained fibrils and can be seen both in the *A* and *I* bands. This periodic variation in density is not as regular as in the case with collagen. However, in certain cases, the distance between the beads or knots is fairly uniform, giving the fibril a fine cross-striated appearance. The axial distance between striations is about 400 Å in stretched fibrils and does not show obvious differences in the *A* and *I* bands. There is some evidence, which is being further investigated, that the distance is less in contracted than in relaxed fibrils.

If muscle fibers are subjected to strong mechanical agitation in a suitable salt solution, they may be dispersed in the form of a suspension showing strong streaming double refraction. When weakly alkaline KCl solutions are employed (Greenstein and Edsall, 1940), one obtains preparations such as have been used to study particle size and molecular weight (Mehl, 1938; Edsall, 1942; Schramm and Weber. 1942).

Electron microscope examination has been made of such myosin preparations from vertebrate and invertebrate muscles. The myosin particles exist as filaments (PLATE 5a) similar to those previously described by Ardenne and Weber (1941). The edges of the filaments are not sharp, but show irregularities which are probably related to the periodic structure appearing in the filaments as observed in intact myofibrils. No density differences, such as occur in the *A* and *I* bands,

were found in isolated filaments, suggesting that, in the extraction of myosin, the A substance is dissociated from the filaments.

PLATE 5b shows distribution curves of the widths and lengths of the isolated filaments for the myosin extracted from rabbit and lobster muscles. The widths range from 50 to 250 Å, corresponding with those of filaments observed in myofibrils. However, the filament lengths show great variation, being mostly less than 15,000 Å. Apparently the filaments, which extend indefinitely in intact myofibrils, are randomly fragmented by the extraction procedure, producing highly polydisperse myosin suspensions. The range of lengths includes those deduced from streaming double refraction experiments (Edsall, 1942). It is obvious that the myosin particles, as conventionally prepared, do not represent a molecular species. Hence, determinations of molecular weight by ultracentrifuge and other methods do not have the significance commonly attributed to them.

MOLLUSCAN SMOOTH MUSCLE

Paramyosin

When molluscan adductor muscles (*Venus, Mya, Mytilus, Ostrea,* and *Pecten*) are macerated in a mortar with 0.3 *M* KCl, a viscous mass is obtained from which may be prepared a suspension of needle-shaped fibrils readily observed with the dark field microscope. This fibrous protein forms a considerable portion of the muscle substance.

The solubility properties of this protein differ from those of myosin. Whereas they are insoluble in KCl solutions in the range of concentration up to about 0.45 *M*, the fibrils disintegrate when placed in Edsall's solution, and are no longer visible in the dark field or electron microscope. The dissolved protein can be reprecipitated by dialysis or addition of water. However, the result is a flocculent precipitate containing no fibrils resembling the original highly structured ones.

To distinguish this protein from myosin, it has been given the name *paramyosin.*

Unstained paramyosin fibrils manifest a mottled appearance when viewed in the electron microscope. However, when the fibrils are appropriately stained, a structure of great geometric regularity is observed (Hall, Jakus, and Schmitt, 1945). The stain is absorbed in transverse bands which have an axial periodicity of 145 Å (PLATE 6a). When the stain is appropriately applied, the transverse bands appear as series of equidistant spots (PLATE 6b). The spots of adjacent bands lie on diagonals which make angles of 28.2° and 38.8° with the axis

(PLATE 6c). The whole structure presents a lattice containing an axial periodicity five times that of the band separation, or $5 \times 145 = 725$ Å. Thus, if a line is drawn through any spot parallel to the axis, other spots on this line will be separated by five bands.

Invertebrate Myosin

If clam adductor muscles are macerated in Edsall's solution (in which the paramyosin fibrils dissolve), fibrous bundles of another type may be obtained by centrifugation. These bundles are composed of fine filaments, closely resembling the myosin filaments of skeletal muscle. Closely adhering to the filaments are dense particles which are frequently aligned to give a cross-striated appearance to the bundles. The axial distance between the striations is about 1100 Å.

Whether this represents a type of cross-striation, in this smooth muscle, which is analogous to that of skeletal muscle, is not clear. However, it should be noted that, in these adductor muscles, which contain a large proportion of paramyosin fibrils, myosin filaments are also present, and the individual filaments resemble those isolated from striated skeletal muscle. Isolation of this material by Edsall's method yields filaments similar to those similarly extracted from vertebrate skeletal muscle, though the length of the particles is significantly shorter.

As yet, no extensive study has been made of the structure of these components in extended and contracted adductor muscles. However, it seems probable that the myosin components are contractile in these muscles, as they are in striated skeletal muscle. No differences in the geometric pattern of the paramyosin fibrils have been observed in fibrils taken from muscles in different physiological states. There is, therefore, no morphological evidence that the paramyosin fibrils actually shorten. Conceivably, they may serve a mechanical function which is related to the ability of these muscles to remain shortened for long periods of time with minimal energy expenditure ("clutch mechanism"). Information is needed about the structural relationship between the paramyosin and the myosin components in the intact muscle fibers. In a variety of invertebrate muscles, the relative proportions of myosin and paramyosin, as estimated by x-ray methods (see below), vary considerably. Thus, in the lantern retractor muscle of *Thyone,* there appears to be no detectable amount of paramyosin, though the collagen content is considerable. Conceivably, this collagen contributes mechanical properties important in the retractor function. Cross-striated myofibrils, on the other hand, appear to contain

no fibrous components within the myofibrils, other than the contractile myosin filaments. Such structural variations may be related to the different physiological properties of the muscles.

While the above is somewhat speculative, it seems useful to call attention to the possibility, suggested in a variety of ways during the last century, that contractile tissue may contain fibrous constituents which do not contribute to the rapid and reversible, twitch-like contraction, but may have a mechanical function of another sort. Further electron microscopic investigation may be expected to throw light on this problem.

X-RAY DIFFRACTION STUDIES

Except for those features which are characteristic of the paramyosin and collagen, when present, the diffraction patterns of a wide variety of striated and smooth muscles are essentially the same and, therefore, have been attributed to the myosin filaments. The differences of structure observed at the light microscope level of resolution are not reflected in the x-ray structure of the myosin components.

Myosin Structure

Most of the x-ray studies of myosin have been concerned with the wide-angle patterns. The literature dealing with this aspect has been reviewed in the papers of Astbury and Dickinson (1940), Astbury (1943), and Schmitt (1944).

The wide-angle patterns are closely similar to those of a number of fibrous proteins, such as keratin and fibrin, and belong to the *alpha* class of Astbury. As such, they reveal little that uniquely reflects the structure of myosin. Information which is more pertinent to the colloidal organization of muscle may be deduced from the small-angle patterns derived from the larger structural periodicities.

Recently, Bear (1945) has obtained evidence regarding a common set of small-angle diffractions observed with a wide variety of muscles, including the adductor muscles of *Mytilus, Venus, Anodonta, Mya,* and *Pecten,* the retractor muscles of *Phascolosoma* and *Thyone,* frog sartorius, and the retractor penis muscle of the dog. *Some* of these muscles also yield small-angle diffractions characteristic of paramyosin (see below) or collagen, but in *all* of them are found the indications of a common fibrous element, presumed to be myosin.

The myosin small-angle pattern is somewhat poorly defined, because of the effects of imperfect orientation, but is nevertheless definitely

recognizable. It consists, principally, of meridional and near-meridional diffractions corresponding to spacings of 58, 51, 272, 18.7, 13.8, 11.2, 9.2 and 6.9 Å. Judging from details, such as concentration of intensity with respect to the meridian, the largest truly meridional diffraction is probably that yielding the 27 Å value. The existence of the closely spaced and probably off-meridional 58 and 51 Å spacings suggests that the true fiber axis period is at least 350 to 420 Å.

MacArthur (1943) earlier had cited the myosin spacings up to the 27 Å value from data of Astbury for frog sartorius, and concluded that the axial repeating period is 658 Å. From the more complete data now available, it is clear that this conclusion is not tenable.

The off-meridional diffractions of myosin have a lateral separation on the patterns corresponding to a transverse spacing roughly estimated to be about 115 Å.

The meridional and near-meridional small-angle diffractions are observed, as quoted, only when the axis of the muscle is perpendicular to the x-ray beam. From a pronounced movement of the diffractions along the meridian and away from the center, which occurs with departure from perpendicular orientation of the specimen, it is concluded that the fibrous diffracting elements are very thin, this dimension being probably of the order of size of the largest meridional spacing, which is 27 Å. This thinness may be important in providing the fibrous elements with the high degree of chemical and structural lability which must characterize these contractile elements. This lability is also in evidence in the fact that the myosin diffractions are generally faint, diffuse, and susceptible of damage by physical and chemical manipulation of the muscle, making it difficult to obtain satisfactory diffractions.

According to Astbury and Dickinson (1940), small amounts of contraction (10 to 20%) cause relatively little change in the wide-angle pattern. More drastic contraction has been associated with the supercontracted state, which cannot be well characterized by direct x-ray evidence. As yet, no data are available concerning the small-angle diffractions in contracted muscles. To obtain such data involves great, though probably not insuperable, technical difficulties. Evidence of this sort may be more informative than the wide-angle data.

Paramyosin Structure

The diffraction patterns of paramyosin fibrils from the adductor muscles of the clams *Mya*, *Venus*, and *Anodonta* have been described by Bear (1944). The wide-angle pattern is that of a typical alpha

protein fiber. However, the small-angle pattern is the most highly developed thus far found for any protein fiber.

The meridional diffractions are orders of a spacing of 145 Å. However, layer-line diffractions indicate that the true fiber axis period is five times this value, or close to 725Å. Since more than forty layer lines have been measured, it would appear that the periodic structure is repeated with great regularity. These patterns have been obtained on extended and dried muscles. Such regularity of structure was unexpected in a fiber having the chemical and mechanical lability of muscle. However, as pointed out above, the function of the paramyosin fibrils is not yet clear. They may prove to have mechanical functions other than that of contraction.

Non-meridional diffractions in the small-angle patterns of clam muscles indicate the presence of a large spacing transverse to the fiber axis. These diffractions are not as well resolved as the axial diffractions, but the spacing is estimated to have a value in the range of 200 to 350 Å.

The correlation between the small-angle x-ray pattern and the electron microscope observations of the structure of paramyosin fibrils is remarkably good. Not only is there exact agreement about the presence of the transverse bands with longitudinal spacing of 145 Å, but the intra-band structure which gives rise to the master axial period of 725 Å, predicted by the x-ray results, has been directly visualized in the electron micrographs.

Comparative X-Ray Studies of Muscle

The myosin small-angle diffractions are the only ones which have been found in the patterns of all types of muscle thus far examined. Indeed, this forms the basis for assigning them to myosin. Even in those clam muscles in which the presence of paramyosin is most evident from the x-ray diagrams, the pattern of myosin is also visible. The small-angle x-ray patterns provide useful information concerning the relative proportions of myosin, paramyosin, and collagen. These are indicated in TABLE 1. Such comparisons may have functional significance, as discussed above.

SUMMARY AND CONCLUSIONS

Electron micrographs of myofibrils of cross-striated muscles reveal a structure resembling that described histologically, but add many important facts. The *A* and *I* bands are distinguished chiefly by the

TABLE 1

RELATIVE INTENSITIES OF THE MYOSIN AND PARAMYOSIN X-RAY DIFFRACTION SYSTEMS EXHIBITED BY VARIOUS MUSCLES

Muscle	Myosin Diffractions	Paramyosin Diffractions
Mytilus adductor	+	+++++
Venus adductor, *w*	++	++++
Anodonta adductor, *w*	++	+++
Anodonta adductor, *t*	+++	+++
Mya adductor	+++	+++
Venus adductor, *t*	++++	++
Pecten adductor, *w*	++	++
Phascolosoma retractor	+	+
Dog retractor penis	+	−
Pecten adductor, *t*	+	−
Thyone retractor	+	−
Frog sartorius	+	−

Rough visual estimates of the relative intensities of the myosin and paramyosin diffractions on the patterns of the various muscles are indicated by +; *w* and *t* refer, respectively, to the "white" and "tinted" components of the muscles which possess them.

presence, in the former, of dense, stainable material which redistributes itself during contraction. This "*A* substance" is closely associated with very thin filaments which make up the bulk of the myofibril; hence, the filaments are presumed to be composed chiefly of myosin. The myosin filaments follow a relatively straight course through *A* and *I* bands. The relative isotropy of the *I* bands is, therefore, not due to random orientation of the filaments, but must be explained in other ways. Nor is there any coiling or folding of the filaments in contracted myofibrils. Contraction must, therefore, involve structural rearrangements of components within the myosin filaments.

Myosin prepared by the method of Greenstein and Edsall consists of a suspension of filaments resembling those in intact myofibrils, but having a wide distribution of lengths, due to random fragmentation during the extraction process. Accordingly, estimates of particle length and weight derived from streaming double refraction and sedimentation measurements on such preparations have no obvious relation to the molecular status of this protein. X-ray data suggest the presence of very thin myosin arrays having thickness of the order of only a few polypeptide chains. This type of structure would make possible exposure of active groups to the chemical environment, alteration of which results in change in configuration, *i.e.*, contraction and relaxation.

Further advance in an understanding of the contractile mechanism from the structural approach requires more detailed information about

the internal organization of the myosin filaments. The resources of the x-ray and electron microscope techniques have by no means been exhausted, and important further advances may be expected. The problem is not insuperable, technically, but requires close correlation of the structural, chemical, and physiological data.

LITERATURE CITED

1. **von Ardenne, M., & H. H. Weber**
 1941. Elektronenmikroskopische Untersuchung des Muskeleiweisskörpers Myosin. Koll. Z. **97**: 322–325.
2. **Astbury, W. T.**
 1943. X-rays and the stoichiometry of the proteins. Adv. Enzymology **3**: 63.
3. **Astbury, W. T., & S. Dickinson**
 1940. X-ray studies of the molecular structure of myosin. Proc. Roy. Soc. London B **129**: 307–332.
4. **Bailey, K.**
 1944. The proteins of skeletal muscle. Adv. Protein Chem. **1**: 289.
5. **Bear, R. S.**
 1945. Small angle x-ray diffraction studies on muscle. J. Am. Chem. Soc. **67**: 1625–1626.
6. **Dubuisson, M.**
 1942. Sur la répartition des ions dans le muscle strié. Arch. Int. Physiol. **52**: 439.
7. **Edsall, J. T.**
 1942. Streaming birefringence and its relation to particle size and shape. Adv. Colloid Sci. **1**: 269–316.
8. **Engelmann, T. W.**
 1875. Contractilität en doppelbrechend Vermogen. Pflüg. Arch. **11**: 432–464.
9. **Greenstein, J. P., & J. T. Edsall**
 1940. The effect of denaturing agents on myosin. I. Sulfhydryl groups as estimated by porphyrindin titration. J. Biol. Chem. **133**: 397–408.
10. **Hall, C. E., M. A. Jakus, & F. O. Schmitt**
 1945. The structure of certain muscle fibrils as revealed by the use of electron stains. J. Appl. Phys. **16**: 459–465.
11. **Hall, C. E., M. A. Jakus, & F. O. Schmitt**
 1946. An investigation of cross striations and myosin filaments in muscle. Biol. Bull. **90**: 32.
12. **Jordan, H. E.**
 1933. The structural changes in striped muscle during contraction. Physiol. Rev. **13**: 301–324.
13. **MacArthur, I.**
 1943. Structure of α-Keratin. Nature **152**: 38–41.
14. **MacCallum, A. B.**
 1905. On the distribution of potassium in animal and vegetable cells. J. Physiol. **32**: 95–128.
 1908. Ergeb. Physiol. **7**: 552.
15. **Mehl, J. W.**
 1938. Double refraction of flow of protein solutions. Cold Spring Harbor Symp. Quant. Biol. **6**: 218-227.
16. **von Muralt, A. L., & J. T. Edsall**
 1930. Studies in the physical chemistry of muscle globulin. III. The anisotropy of myosin and the angle of isocline of flow. J. Biol. Chem. **89**: 315–386.

17. **Richards, A. G., Jr., T. F. Anderson, & R. T. Hance**
1942. A microtome sectioning technique for electron microscopy illustrated with sections of striated muscle. Proc. Soc. Exp. Biol. & Med. **51**: 148–152.

18. **Schmidt, W. J.**
1937. Die Doppelbrechung von Karyoplasma, Zytoplasma und Metaplasma. Bornträger. Berlin.

19. **Schmitt, F. O.**
1944. Structural proteins of cells and tissues. Adv. Protein Chem. **1**: 25–68.

20. **Schramm, G., & H. H. Weber**
1942. Über monodisperse Myosinlösungen. Kolloid Z. **100**: 242–247.

21. **Scott, G. H.**
1932. Distribution of mineral ash in striated muscle cells. Proc. Soc. Exp. Biol. & Med. **29**: 349–351.

22. **Scott, G. H., & D. M. Packer**
1939. Electron microscope study of magnesium and calcium in striated muscle. Anat. Rec. **74**: 31–45.

23. **Sjöstrand, F.**
1943. Eine neue Methode zur Herstellung sehr dünner Objektschnitte für die elektronenmikroskopische Untersuchung von Geweben nebst einigen vorläufigen elektronenmikroskopischen Beobachtungen über den submikroskopischen Bau der Skelettmuskelfaser. Arch. Zool. **35A**: 1.

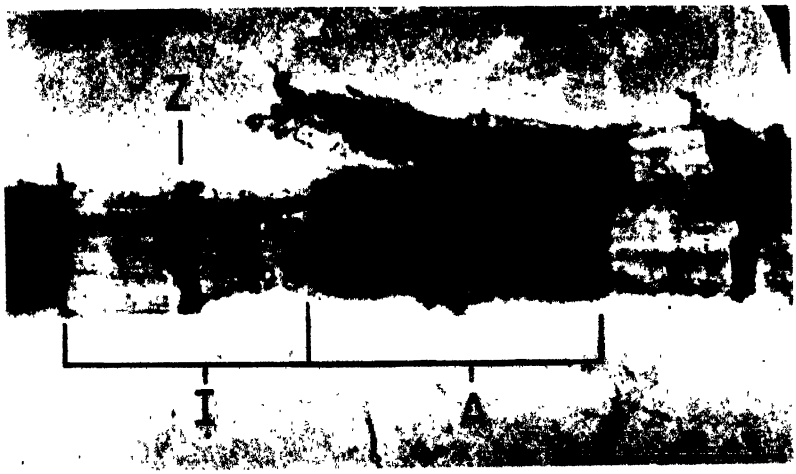
24. **Smith, E. C. B.**
1937. Native and denatured muscle proteins. Proc. Roy. Soc. London B **124**: 136–150.

25. **Winter, L. B., & W. Smith**
1922. Formation and distribution of phosphates in muscle. J. Physiol. **56**: 227.

PLATE 4

a. Myofibril from frog sartorius, stretched about 30 per cent, fixed in formalin, stained with phosphotungstic acid; x 30,000. (Reprinted from Biol. Bull.[11])

b. Myofibrils from frog sartorius contracted, fixed in formalin, stained with phosphotungstic acid; x 30,000. (Biol. Bull.[11])



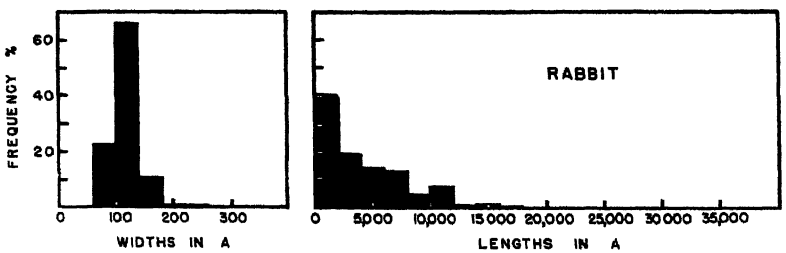
a



b



a


FREQUENCY %
60
40
20
0
100
200
300
WIDTHS IN A
RABBIT
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
LENGTHS IN A

b

PLATE 5

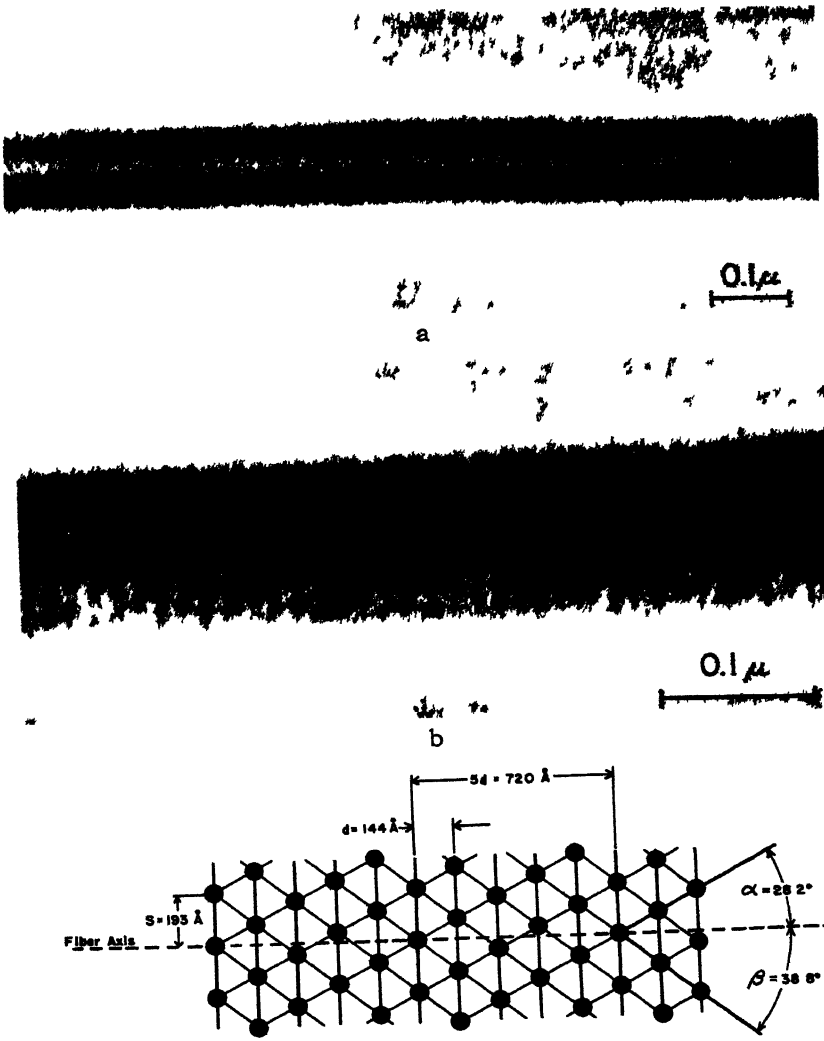
a Myosin filaments from rabbit leg muscle × 35 000 (J Biol)
b Distribution of lengths and widths of myosin filaments from rabbit leg muscle (J Biol)

PLATE 6

a Paramyosin fibrils stained with phosphotungstic acid, showing 145 Å spacing, × 100 000 (J Appl Phys[20])

b Paramyosin fibrils stained with phosphotungstic acid, showing lattice structure, × 210 000 (J Appl Phys[20])

c Diagrammatic lattice showing geometrical relations and dimensions between stained regions in paramyosin fibrils (J Appl Phys[20])


0.1μ
a
0.1 μ
b
5d = 720 Å
d = 144 Å
S = 193 Å
Fiber Axis
α = 28.2°
β = 38.8°

MUSCULAR CONTRACTION

PART III

CHEMISTRY

THE MAIN CHEMICAL PHASES OF THE RECOVERY OF MUSCLE

By Otto Meyerhof

Department of Physiological Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania

Our knowledge in this field has been rather well established for the last few years, and little can be added that is really new. Some facts have received additional light by new methods, especially by isolation and purification of single enzymes and by the use of labelled radioactive phosphate. These methods have mainly helped to confirm what was already known of the metabolism of the active muscle. But we should be more critical in regard to the claim of completeness of this picture of the recovery process. Quantitative disagreements in the application of some of these methods prove that several important steps are still unknown to us.

Nevertheless, I think our present-day picture of the chain of events in the recovering muscle to be essentially correct. Every single step and relationship was made a matter of controversy during recent years, but, in the end, the older interpretation was confirmed. The anaerobic nature of the fundamental chemical process of contraction, as found by A. V. Hill, was questioned by Dr. Sacks.[1, 2] The same author challenged the view that lactic acid was formed *via* phosphorylated intermediaries, and that phosphocreatine (PC) and adenosinetriphosphate (ATP) intervene in the active muscle in the same manner as in enzymatic extracts.[3] Such criticism deserves our full appreciation, even if it proves to be unjustified, because it keeps the research worker from relapsing into complacency. It stimulates him to devise new schemes for confirming his findings and interpretations. This also happened in our case.

The time-relation of oxygen consumption to heat-production was taken up by David K. Hill,[4] son of Professor A. V. Hill. He found complete coincidence of oxygen consumption and oxidative heat in isolated muscles at 0° C., and confirmed the anaerobic nature of the contraction process beyond any doubt.

Figure 1 shows the special features of the volumetric respiration chamber, used by D. K. Hill (all measurements are made at the temperature of melting ice). The resting respiration is cancelled out by using the symmetrical frog sartorius, which is not stimulated, in one

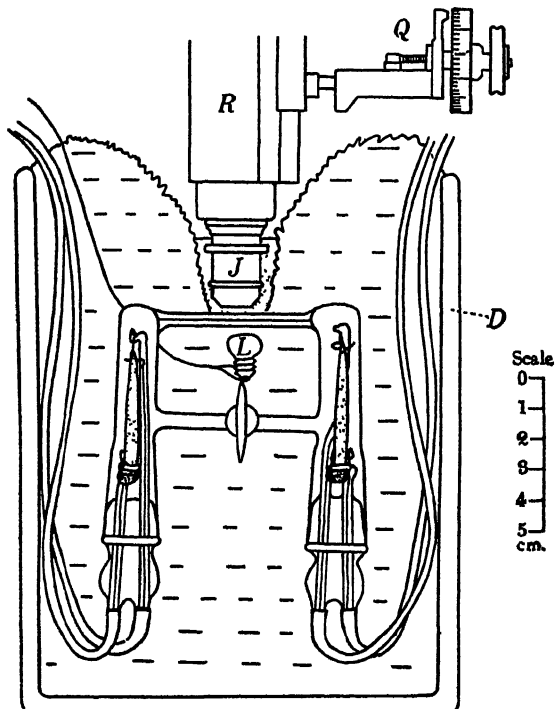


FIGURE 1. Volumeter in ice. *D*, wall of Dewar flask; *L*, flash-lamp bulb; *J*, objective of microscope lowered into water; *R*, tube of microscope; *Q*, micrometer screw.

side of the differential apparatus. CO_2 in both chambers is absorbed by Na_2CO_3 spread on filter paper. The excess of oxygen consumption on the stimulated side is observed by the shift of an index drop in a capillary, connecting both chambers, with the help of a microscope. FIGURE 2 shows the time-course of the oxygen consumption: *A* is the curve obtained after correction for the heat-expansion, *B*, after correcting for the slow diffusion of oxygen into the muscle, while *C* is the theoretical curve which would be expected if the oxygen consumption coincided with contraction and initial heat.

In FIGURE 3, the correspondence of oxidative heat and oxygen consumption is shown: *A* and *B* for pH 7.2, *C* and *D* for pH 6. Oxygen consumption is considerably slowed down with decreasing pH. Most important, however, is the close parallelism of heat-production and oxygen consumption, because it demonstrates that the endothermic

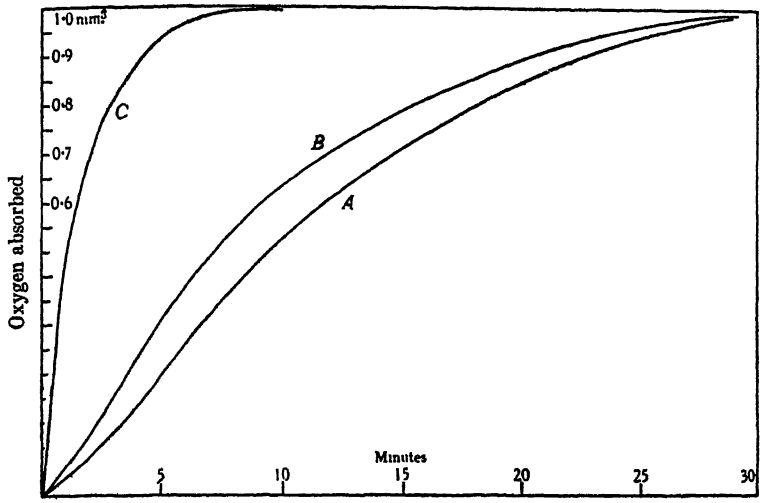


FIGURE 2. *A*, the record uncorrected except for initial heat. *B*, after correction for diffusion. The muscle was of thickness 0.8 mm. *C*, the curve which would be followed by the index if consumption of oxygen coincided with the activity.

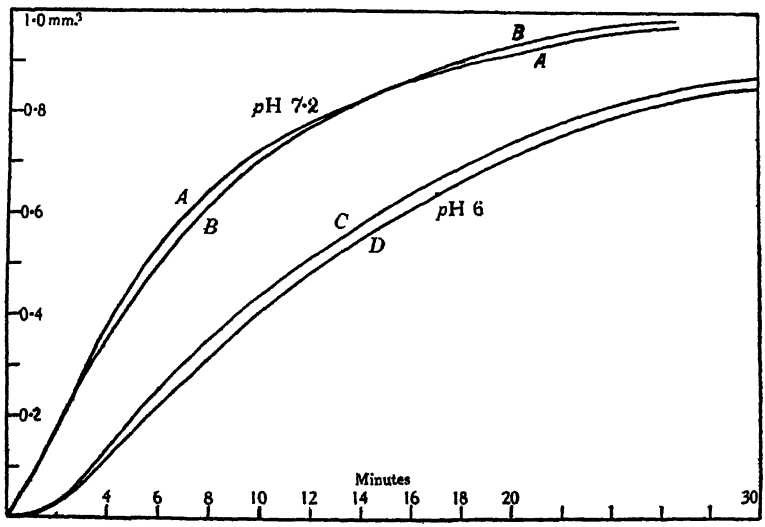


FIGURE 3. Comparison of the time courses of oxygen consumption and of heat production. At pH 7.2. *A*, oxygen; *B*, heat. At pH 6.0: *C*, heat; *D*, oxygen.

syntheses, coupled with the oxidation, remain throughout the whole time of recovery in fixed proportion to the oxidation, actually consuming 50 per cent of the oxidative heat.[5]

The speed of the oxidative recovery is not determined by physical

factors like diffusion of oxygen, but by chemical enzymatic conditions of the muscle. According to D. K. Hill, the content of cytochrome oxidase decides the speed. Insect muscles, like those of the locust with a high content of cytochrome and cytochrome oxidase, show at the same temperature a rate of recovery oxidation about 100 times that of frog muscle with low content of cytochrome oxidase. Moreover, sodium azide, which, according to Stannard,[6] in higher concentrations inhibits the recovery oxidation completely, but not the resting respiration of muscle, slows the recovery oxidation down in the range of 2×10^{-5} to 2×10^{-4} M. This, too, must be explained by inhibition of the cytochrome oxidase.

The second objection of Dr. Sacks was based on experiments with radioactive phosphate.[3] The author did not find such an exchange of labelled phosphate between the inorganic fraction and PC and ATP as he expected on the basis of the enzymatic exchange reactions. He assumed, therefore, that these exchange reactions did not take place in the active muscle. Dr. Herman Kalckar[7] cleared up this contradiction with his usual perspicacity and precision. He showed that Dr. Sacks overlooked the difference of extracellular and intracellular inorganic phosphate. Only the latter can be expected to be in free enzymatic exchange with the phosphate esters, since the muscle cellular membranes are very little permeable to phosphate. Since the inorganic labelled phosphate is added from the outside, generally from the circulating blood, there is ten to a hundred times more extracellular than intracellular labelled phosphate. In Kalckar's own experiments, 5 to 7 per cent of the total labelled inorganic phosphate was intracellular. In all experiments done by Sacks and by other authors before Kalckar's investigation, the exchange between intracellular labelled inorganic P and PC and ATP was already complete in the shortest experimental interval which they applied, and, therefore, it could not be enhanced by periods of activity. In Kalckar's own experiments with resting muscle, more than half of the intracellular radioactive phosphate had exchanged between the inorganic phosphate and the energy-rich phosphate groups in about twenty minutes. Experiments with active muscle had to be done in much shorter time intervals, to find out how quick the isotope equilibrium would be approached between the inorganic phosphate and the energy-rich phosphate groups. We found, in 1938 in Heidelberg, with enzymatic extracts,[8] that the exchange of biological concentrations of ATP with inorganic phosphate in the presence of the complete enzyme system was finished in about 50 seconds with a ten times diluted enzyme. With undiluted enzyme, it

would have been completed in a few seconds. Although we do not know what factors may intervene in the living muscle, and the common pool of exchangeable phosphate with PC, ATP, and phosphopyruvic acid must be considered, exchange in the active muscle may be completed in less than one minute.

I think that there cannot be the slightest doubt that glycolysis in the active, as well as in the resting, muscle passes through the phosphorylated intermediaries known from enzymatic studies, and that PC breakdown and synthesis are connected with it by way of phosphate transfer from ATP and ADP. Glycolysis, here, means strictly formation of lactic acid and not oxidation of sugar. Dr. Ochoa will discuss the role of phosphorylation in oxidation. Some authors assume by implication that lactic acid formation is always the first step of sugar oxidation. This is unwarranted and, for the steady state, highly improbable, because, in the presence of oxygen and of oxidative catalysts, pyruvic acid would not be reduced to lactic acid. Moreover, the pathway of sugar oxidation *via* pyruvic acid is not the only possible one, although probably the main route. Furthermore, according to the R. Q. of about .95, carbohydrate is the prevailing, but not the exclusive, fuel of the active muscle in the presence of oxygen.

New viewpoints on the question as to how the oxidation of fat can participate in the recovery process and transfer its energy to the synthesis of high energy phosphate bonds arise from the recent findings of Breusch, Weinhouse, Lehninger, and others. According to Breusch,[9, 10] two carbon residues of fatty acids, formed by beta-oxidation *via* acetoacetic acid, condense with oxalacetic acid to citric acid. The further oxidation of citric acid would then be coupled with endergonic synthesis in the same way as in the oxidation and degradation of sugar by way of the Krebs cycle. According to the important discovery of Lehninger,[11] catalytic amounts of ATP are necessary for the oxidation of fatty acids in the liver. This means that labile phosphate must be transferred to phosphorylated intermediaries, undoubtedly acylphosphates, and that these, in turn, would generate new energy-rich phosphate. If this enzymatic process plays a larger role in the liver than in the muscle, as seems to be the case, then it would be more important for the recovery of the muscle *in situ* than for the isolated muscle. It would explain that, in the whole animal as in man, work could be done nearly as efficiently by oxidation of fat as of carbohydrate. Synthesis of lactic acid to glycogen could be easily coupled with the oxidation of fat. In any case, generation of energy-rich phosphate is always the first step of synthesis.

The interconnection of the single phases of recovery is shown in the following diagram (FIGURE 4) by Dr. Fenn,[12] who has pictured, in a diagrammatic drawing, the abstract scheme of consecutive coupled reactions, formerly sketched by Dr. Lohmann and myself (taken from

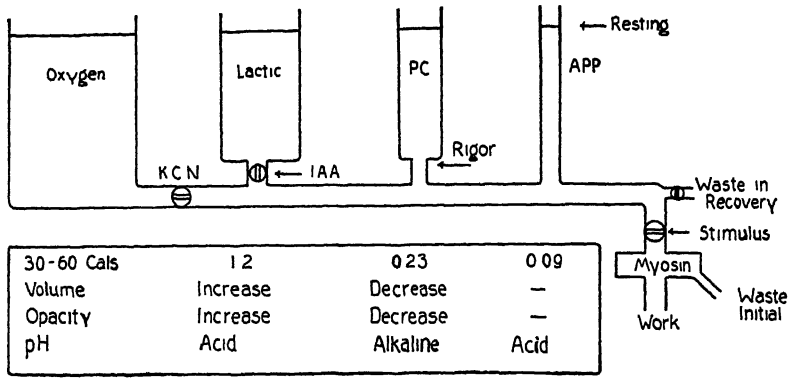


FIGURE 4 Diagram representing various energy reservoirs in muscle Energy is derived from oxidation, glycolysis (lactic), phosphocreatine breakdown (PC), or adenylpyrophosphate breakdown (APP)

Professor Höber's book, *Physical Chemistry of Cells and Tissues*). The energy reservoirs, as contained in one gram of living frog muscle, are pictured by the diameter of the four columns of communicating liquid which all are able to drive the myosin-engine of the muscle. Oxygen means the total amount of available oxidative material, especially carbohydrate; lactic acid means the amount of glycogen convertible into lactic acid by stimulation, until complete fatigue. The points of attack of KCN and iodoacetic acid, as well as the hypothetical site of the stimulus, are also shown. Nevertheless, I will give our primitive scheme[13] here again (FIGURE 5), because, in these coupled reactions, not only the total amount of available energy is important, but the quantitative relations on the basis of stoichiometric coupling.

I discussed this question more thoroughly two years ago,[14] and will come back to it only so far as the correlation is concerned between the chemical processes and the phases of heat, either measured by thermocouples or in calorimeters. Generally, I propose to replace the former notion of caloric quotient of lactic acid by the caloric quotient of high energy phosphate, and express this, not in cals per gram but in cals per mole. For this calculation, the lactic acid formed is equivalent to two energy-rich phosphate bonds generated by the formation of one mole lactic acid, one from 1, 3 diphosphoglyceric acid, and one from phospho-

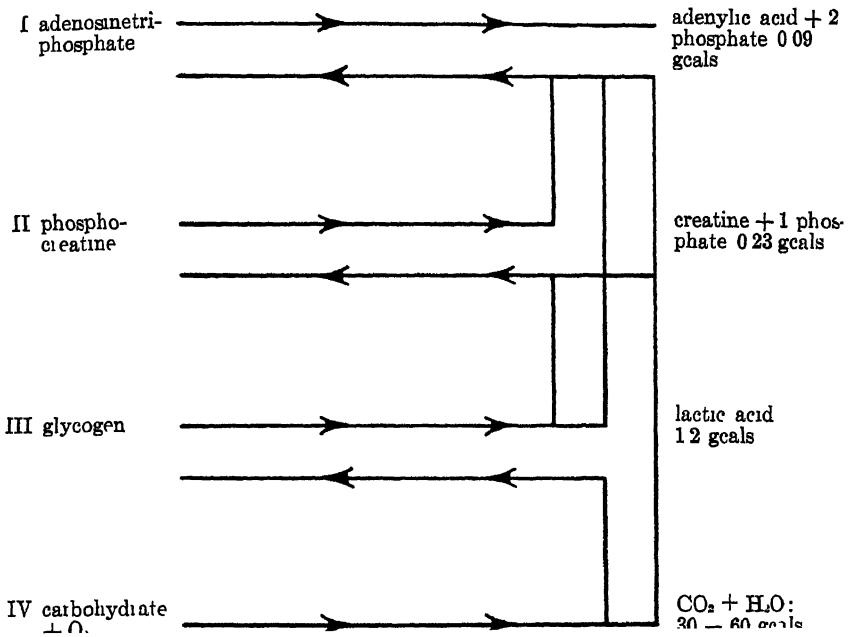


FIGURE 5

pyruvic acid. By the use of this new term, we are independent of what really happens in the muscle at every moment, whether the reaction measured by the heat is decomposition of ATP to ADP or to adenylic acid, or splitting of PC with the catalytic help of ATP, or formation of lactic acid alone, or any sum of these three processes. In every instance, we have to compare the total balance of changes with the heat. The caloric quotient of high energy phosphate per mole would be around 12,000 small cals., according to the heat measured during enzymatic splitting. In the living muscle, it is somewhat higher, partly (to the extent that new hydrogen ions are formed) by additional heat from neutralization of protein, partly from unknown concomitant reactions, and may generally amount to about 13,000 to 14,000 cals.[15]

A closer correlation of the single phases of heat with the underlying chemical reactions is much more hypothetical. When we take our present-day picture as a guide, most of the single chemical reactions would cancel each other out regarding the heat, and the resulting coupled reactions would be nearly thermoneutral, for instance, splitting of creatinephosphate with resynthesis of ATP from ADP. The only two instances of uncompensated positive heat are the first initial

splitting of ATP by the adenylpyrophosphatase (apyrase) and the delayed oxidative recovery process, with an outburst of heat amounting to 50 per cent of the calculated combustion heat of carbohydrate. As I stated two years ago, this amount of oxidative heat corresponds to 5 energy-rich phosphate bonds resynthesized for every mole of oxygen consumed. In this picture, the initial heat of contraction, as measured by Hartree and A. V. Hill, would be explained by splitting of ATP.

There remains still another small phase of heat, the so-called delayed anaerobic heat, which was carefully analyzed by D. K. Hill in 1940.[16] On the average, it amounts to 10 per cent of the initial heat. Its interpretation is difficult, because of its inconstancy. According to David Hill, it is absent in the first contractions of a completely fresh muscle, except for a small negative phase at the onset, and becomes gradually more pronounced with fatigue of the muscle. It is more regular and better defined in the alactacid contraction of a muscle poisoned with iodoacetic acid. FIGURE 6 is taken from the paper of D. K. Hill. *A* corresponds to the anaerobic delayed heat of a normal moderately fatigued muscle, *B* to the delayed heat of a poisoned muscle in the same state of fatigue. Since the small endothermic phase in the beginning happens in both cases, it cannot be explained by the voluntary endothermic coupled reaction of phosphate transfer from phosphopyruvic acid to creatine. It may be caused, as Hill suggests, by

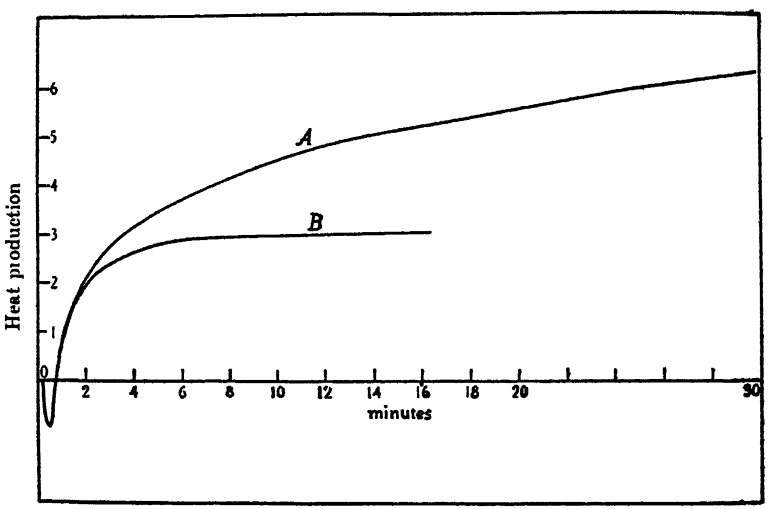


FIGURE 6. *A*, anaerobic delayed heat at pH 7.2. *B*, a member of a series under "alactacid" conditions for comparison. Duration of tetanus 12 sec. The heat-production is expressed as percentage of initial heat.

the slightly endothermic phosphate transfer from PC to adenylic acid or to ADP. Indeed, the splitting of PC gives only 11,000 cals., while the splitting of ATP gives 12,000 cals. per mole phosphate. This difference is not very great, but seems to be well established. For PC, a great number of measurements were done in my former work at the Heidelberg Institute with very pure preparations, and gave consistently the smaller figure of 11,000 cals.[17] The less exact heat measurements on ATP, which gave 12,000 cals. per mole phosphate,[18] were recently repeated by Dr. Ohlmeyer in Tübingen, using a new and very exact method for measuring the heat, and were confirmed by him.[19]

We may ascribe the positive delayed anaerobic heat phase to an excess of PC split over synthesis of ATP by intervention of the apyrase, and the extra amount of ill-defined heat in the unpoisoned muscle (FIGURE 6, Curve *A*) to an excess of lactic acid formation over the resynthesis of two moles PC per mole lactic acid, by which the heat would be compensated. McKeen Cattell and Einar Lundsgaard showed, in 1933,[20] that the efficiency of a muscle working in oxygen is about 15 per cent less in a long series of about 100 twitches than in a short series of only 20 twitches. They even then ascribed this difference to the extra amount of delayed anaerobic heat in higher degrees of fatigue of the muscle.

The sequence of the recovery processes in muscle was elucidated by several independent techniques. One was the study of the single reaction steps and their coupling in enzymatic extracts. Other approaches were by means of the concomitant physical changes which accompany fatigue and recovery and which can be used as indicators of the underlying chemical reactions. Perhaps the most important tool was the poisoning of the muscle with iodoacetic acid, which, according to Lundsgaard's discovery, produces the alactacid contractions.[21] The behavior of the muscle poisoned with iodoacetic acid is so well known that I will mention only some main features of this altered condition of the working muscle. The immediate consequence of the complete inhibition of lactic acid formation is accumulation of hexosemono-and-diphosphate and the decomposition of PC in an enhanced rate. After the bulk of PC is split, the muscle becomes inexcitable, its ATP content diminishes rapidly, and it goes into rigor. The inhibiting effect itself is caused, undoubtedly, by the oxidation of sulfhydryl groups in the enzyme protein of the triosephosphate-oxidase. If this oxidizing enzyme is blocked, glyceraldehyde-3-phosphate is no longer oxidized by cozymase to 1, 3 diphosphoglyceric acid, no energy-rich phosphate is formed, and, therefore, no PC can be resynthesized

anaerobically. The anaerobic work of the muscle becomes now strictly proportional to the splitting of PC. If all of the PC is split, no more energy-rich phosphate is available to be transferred to the adenylic system, and the dephosphorylation of ATP is no more opposed by synthesis. The rigor which sets in under these conditions could be attributed to the exhaustion either of PC or of ATP. This disappearance of energy-rich phosphate in the muscle is, moreover, the common denominator of all kinds of rigor, *rigor mortis*, heat rigor, chloroform rigor, caffeine rigor, etc. From the recent work of Szent-Györgyi,[22] it appears that the disappearance of ATP is responsible for the rigor, since the protein mixture which he calls actomysin becomes insoluble after decomposition of ATP, and the tension development in rigor is said to be strictly proportional to the diminution of ATP in muscle.

The accumulation of the hexosephosphates must be attributed partly to the damming backwards of the triosephosphates which do not react further. This explains the accumulation of hexosediphosphate. Hexosemonophosphate may also increase by the higher concentration of inorganic phosphate, which drives the reaction of Cori to increased phospholysis of glycogen. Glucose-1-phosphate mutates to the Embden-ester, according to the equilibrium constant.

I may say a few words about the physical changes occurring during recovery. Von Muralt and E. v. Bacyer[23] found in Heidelberg, in 1934, that a muscle in Ringer solution becomes, at first, more transparent to light during fatigue and, in the final stage, more opaque. The increase in transparency goes parallel with the dephosphorylation of PC. Buchthal and Knappeis were able to extend these results by applying them to single muscle fibers.[24] This effect of metabolic changes in increasing the scattering of light by the muscle structures is very interesting, from a physical point of view. However, the change of opacity cannot be evaluated in terms of chemical turnover; therefore, I will not go into more details.

In this respect, the measurement of the volume change is more fertile. We are here on firm ground, because we know, from the study of the enzymatic changes in dilatometers, that the dephosphorylation of ATP gives a volume constriction of 10 cc. per mole phosphate split off, dephosphorylation of PC gives 11.5 cc. constriction, lactic acid formation from glycogen gives 24 cc. dilatation per mole, and so on, on account of the difference in molecular volume of the disappearing and the newly formed substances. The volume constriction during the muscle contraction was first discovered by the Hungarian physiologist

Ernst,[25] but the irreversible changes and their connection with muscle metabolism were first observed in the Heidelberg Institute.[26] A typical curve accompanying a series of isometric tetani is shown in FIGURE 7.[27] The uppermost line, with the roman numbers I, II, III, etc.,

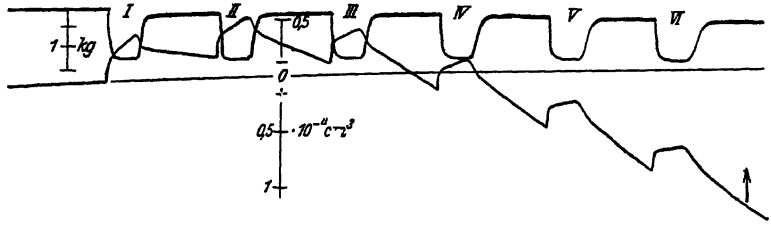


FIGURE 7.

shows the tension development in consecutive two-second tetani, the next line gives the volume change of the muscle imbedded in paraffin oil. Volume contraction is drawn upwards. The sharp constriction during the onset of tension and the lesser augmentation during the maintenance of tension must be attributed to the splitting of ATP and PC, while the slow dilating reaction between the tetani which more and more compensates and finally exceeds the constricting reactions must be attributed to lactic acid formation (after correction for heat-dilatation). This becomes especially obvious by comparison with the volume change in the iodoacetic acid poisoned muscle (FIGURE 8). In FIGURE 8, the poisoned muscle is stimulated to exhaustion with two second tetani. In the first three tetani, the tension development is about the same as in the case of the unpoisoned muscle (FIGURE 7). The curve of volume constriction is, at first, similar, although the total amount of constriction during the tetanus is considerably larger than in the unpoisoned muscle. Still more conspicuous is the absence of the dilating after-effect which, in the other case, soon exceeds the constriction. On the contrary, the constriction as produced by the tetanus remains nearly unaltered, until the next tetanus gives an additional constriction which adds to the first, and so on, until the muscle becomes unexcitable. Since the amount of heat produced by the tetani is the same in the poisoned and unpoisoned muscle, the difference has to be attributed to the greater breakdown of PC and to the complete absence of lactic acid formation in the poisoned muscle.

Nevertheless, this is not the complete story. Firstly, there is some reversible effect of tension development on volume change which is not explained by chemical turnover. This effect may sometimes

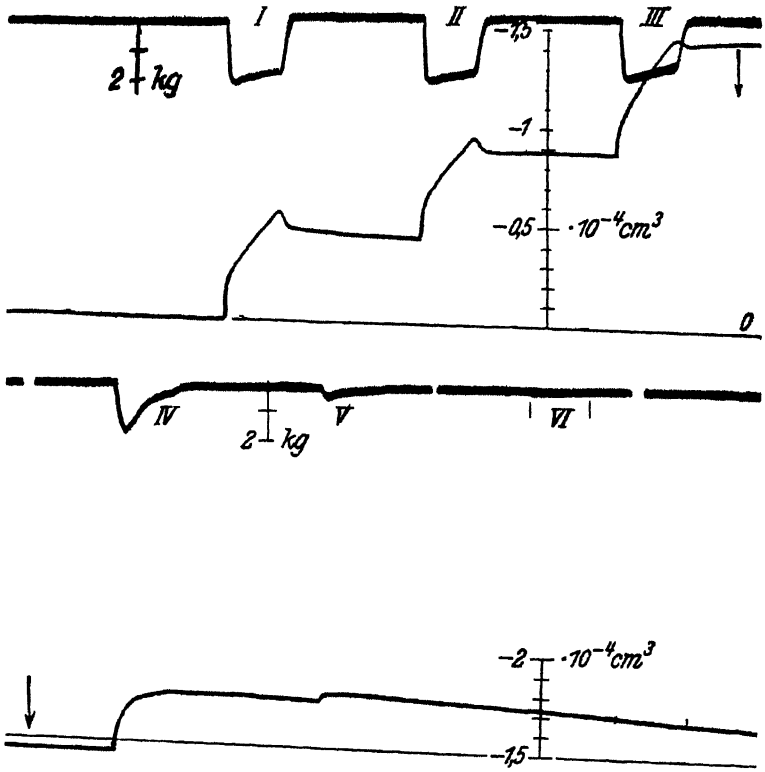


FIGURE 8

change its sign: *e g*, Ernst Fischer has described a sudden increase in volume of sartorius muscle contracting under high tension.[28] Superimposed effects can also be found by applying a strong pull to the gastrocnemius while contracting. These features may be ascribed to physical factors. Moreover, in the summation of constriction of the fatigued muscle, the agreement between the measured volume change and that calculated from the change of molecular volume of the reacting metabolites is unsatisfactory. An unknown factor is involved which gives additional constriction; this being true for the normal fatigued muscle, for the poisoned muscle, and for different forms of rigor. This is brought to light by using a different method of recording with less magnification and a shorter time axis[29] (demonstrated in PLATE 7: *a* and *b* for normal muscles, and *c* and *d* for poisoned muscles; *a* and *c* in Ringer solution, *b* and *d* in paraffin oil). Since recording here is done by transmitted light, the direction of

change is reversed: downward movement means constriction, upward movement means dilatation. The general picture, as found with the other method, is confirmed: total fatigue of normal muscles gives a remaining dilatation (*-M-*) which is a little larger in paraffin oil than in Ringer, while the transient dilatation caused by heat (*-W-*) is very much larger in paraffin oil. Total fatigue of iodoacetic acid-poisoned muscle gives a persistent constriction, somewhat larger in Ringer than in paraffin oil. But there is no quantitative agreement with the chemical changes measured simultaneously: in the normal muscle, as well as in a muscle brought into rigor by heat, the dilatation should be larger, to about 2 to 3×10^{-4} cc. per gram of muscle, than actually observed. (Observed dilatation 2×10^{-4} cc per gram, calculated 5×10^{-4} cc. per gram.) In the poisoned muscle, 5×10^{-4} cc. constriction is observed, but only 3 to 4×10^{-4} cc. are expected from the chemical turnover. An unexplained constriction of $2–3 \times 10^{-4}$ cc. intervenes in all cases and must be attributed to an undiscovered chemical reaction or to a volume change of protein caused by the known reactions.

On the other hand, measurement and calculation of the change of pH during anaerobic fatigue show a satisfactory agreement, as was shown, in 1938, by M. Dubuisson and W. Schulz in the Heidelberg Institute. As was first observed by F. Lipmann and myself with a manometric technique, there is, at first, a shift to alkalinity due to dephosphorylation of PC, which in the later stages of fatigue is followed by a shift to the acid side due to the prevailing formation of lactic acid.[31] In the muscle poisoned with iodoacetic acid, the acidity phase is completely absent. These findings were corroborated and refined by the work of Marcel Dubuisson, using a glass electrode. Furthermore, the work of Dubuisson brought to light two other important phases which precede the alkalinity due to PC breakdown.[32] These preliminary phases can be measured only by a special glass electrode, constructed in such a way that the action potential of the muscle does not interfere with the measurement. The curve of pH change during isotonic tetanic contraction is shown in FIGURE 9: alkalinity is upwards, acidity downwards. The alkalinity phase, *c*, is the already known phase of PC-breakdown followed by the acidity phase, *d*, of lactic acid formation. However, during the contraction itself (represented by the dotted line), there occurs a small acidity shift, *b*, which is attributed to the splitting of ATP to ADP and phosphate, whereby a secondary valence of phosphate is set free with an acid pK of 4.5. In fresh muscles contracting isometrically, this small acidity phase is preceded by a sharp peak of alkalinity coinciding with the onset of

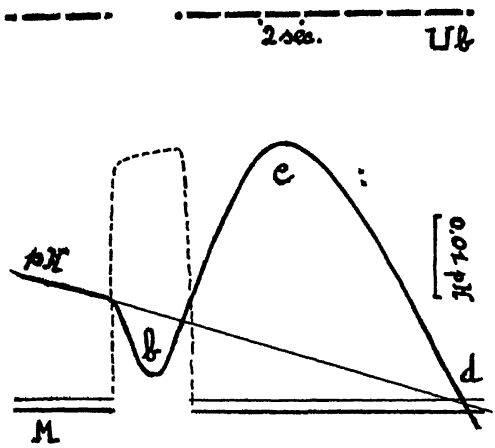


FIGURE 9.

tension. FIGURE 10 shows such a curve and how it is interpreted according to Dubuisson (dotted curves).

That the small acidity change, during contraction, is due to splitting of ATP, is undoubtedly an attractive hypothesis. We should not, however, treat this hypothesis, which was first formulated by K. Lohmann and myself on the basis of enzymatic coupled reactions and theoretical arguments that ATP breakdown initiates the chain of events in activity, as an established fact. We should not be deceived by alluring probabilities and accept them as certainties. To my knowledge, nothing besides the earlier arguments of Lohmann was discovered which adds to the evidence of such a sequence of reactions, and we should treat the problem, accordingly, with caution.

The alkalinity which is only observed in fresh muscles during the onset of tension is of great interest. The connection with tension is established by the fact that passive stretch of resting muscles also gives a similar alkalinity. According to Dubuisson's measurements, the shift is about pH 0.2 to the alkaline side.[33] This effect was first discovered by Margaria,[34] who stained the muscles with suitable indicator dyes, brom cresol purple or brom thymol blue, and observed a color change when the muscles were stimulated in an atmosphere of N_2 with 5% CO_2. The pH shift is too great to be attributed to a stoichiometric reaction of the phosphate esters concerned. Dubuisson assumes a change of the isoelectric point of the myosin fibers, on account of stretching. The phenomenon may be related to the observation of Szent-Györgyi, that the mixing of solutions of myosin and "actin" near the neutral

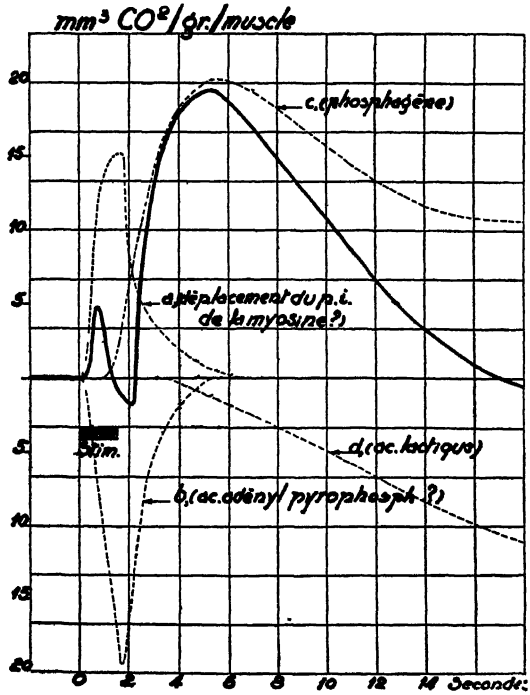


FIGURE 10

point produces a shift to the alkaline side with simultaneous formation of "actomyosin."[22]

Let me add some words to the ideas discussed in recent years in connection with the announcement of Engelhardt and Ljubimova that adenosinetriphosphatase is identical with myosin.[15] That an enzymatic activity is bound to the myosin, as it is usually prepared according to Edsall, is certain, and can easily be demonstrated. Moreover, the same is true for the crystalline myosin prepared by Szent-Györgyi. But I think that the interpretation given to these observations is wrong, that is, the interpretation that the enzyme ATPase is identical with the myosin-fraction, or with about 50 per cent of the muscle protein. Indeed, so far, only the inability of many authors to effect a separation has given support to the claim of Engelhardt. Kalckar[16] has demonstrated that "apyrase" from potatoes is easily adsorbed on myosin, and has suggested that the same may be true for the muscle enzyme. During the last two years, B. D. Polis, working in my laboratory, has tried many different methods of fractionation

by adsorption. At last, we found that, by use of lanthanum nitrate in a weakly alkaline medium, we could precipitate most of the myosin. In the presence of bicarbonate, ATP, and KCN, we then eluted several fractions which exhibit an enzymatic activity two to three times the maximal activity of the myosin before adsorption. If we express the activity by *Qp* (cmm. H_3PO_4 split off per hour at 38°, per mg. dry weight of protein), the activity in the bulk of the myosin is generally *Qp* 2000 to 3000, and *Qp* after elution 5000 to 7000. In exceptional cases, where the activity is, from the start, about *Qp* 5000, it can be raised by adsorption and elution to 15,000.[37] * This is true as well for the myosin of Edsall as for the crystalline myosin of Szent-Györgyi. This fractionation represents only a partial separation, but it shows that the enzyme is only a small part of the myosin or is adsorbed on it.

The final question as to how the ATP reacts with myosin, and how such a reaction is connected with the mechanical changes during muscle contraction, is not a subject of this paper. These questions will be dealt with in the following papers. The only point I would like to raise is that the chemical phase which utilizes the high energy of the phosphate bond must coincide with or precede the onset of the contraction. This conclusion can be inferred from the heat measurements of A. V. Hill.[40] According to these myothermic results, practically no heat is developed during relaxation, unless elastic energy of tension is dissipated. Therefore, even if relaxation is an active process and not only the disappearance of the state of contraction, very little change of energy is connected with it.

BIBLIOGRAPHY

1. **Sacks, J.**
1938. Am. J. Physiol. **122**: 215.
2. **Sacks, J.**
1941. Physiol. Rev. **21**: 217.
3. **Sacks, J.**
1940. Am. J. Physiol. **129**: 227.
4. **Hill, David K.**
1940. J. Physiol. **98**: 207.
5. **Hill, David K.**
1940. J. Physiol. **98**: 454.
6. **Stannard, J. N.**
1939. Am. J. Physiol. **126**: 196.
7. **Kalckar, H. M., Jean Dehlinger, & A. Mehler**
1944. J. Biol. Chem. **154**: 275.

* In discussion, Drs. W. Price & C. F. Cori reported experiments according to which they could separate the ATPase from myosin.[38] See, however, ref. [39].

8. **Meyerhof, O., P. Ohlmeyer, W. Gentner, & H. Maier-Leibnitz**
1938. Biochem. Z. **298**: 396.

9. **Breusch, F. L.**
1943. Science **97**: 490.

10. **Breusch, F. L., & Halit Keskin**
1944. Enzymologia **11**: 243.

11. **Lehninger, A.**
1945. J. Biol. Chem. **161**: 432.
1946. J. Biol. Chem. **162**: 333.

12. **Fenn, W. O.**
1945. In: **R. Hoeber.** Physical Chemistry of Cells and Tissues: 445. Blakiston Co. Philadelphia.

13. **Lohmann, K.**
1937. Angewandte Chemie **50**: 97.

14. **Meyerhof, O.**
1944. Ann. N. Y. Acad. Sci. **45**(9): 377.

15. **Meyerhof, O., E. Lundsgaard, & H. Blaschko**
1931. Biochem. Z. **236**: 326.

16. **Hill, David K.**
1940. J. Physiol. **98**: 460.

17. **Meyerhof, O., W. Schulz, & Ph. Schuster**
1937. Biochem. Z. **293**: 309.

18. **Meyerhof, O., & K. Lohmann**
1932. Biochem. Z. **253**: 431.

19. **Ohlmeyer, P.**
1944. Z. Physiol. Chem. **282**: 38.

20. **Cattell, J. McKeen, & E. Lundsgaard**
1933. J. Physiol. **78**: 247.

21. **Lundsgaard, E.**
1930. Biochem. Z. **217**: 162.

22. **Szent-Györgyi, A.**
1945. Acta Physiol. Scand. **9**(Suppl. 25).

23. **von Baeyer, E., & A. von Muralt**
1934. Pflüg. Arch. **234**: 233.

24. **Buchthal, F., & G. G. Knappeis**
1939. Scand. Arch. Physiol. **82**: 225; **83**: 281.

25. **Ernst, E.**
1925. Pflüg. Arch. **209**: 613.

26. **Meyerhof, O., & W. Möhle**
1933. Biochem. Z. **260**: 454, 469.

27. **Hartmann, H.**
1934. Biochem. Z. **270**: 164.

28. **Fischer, E.**
1941. Biol. Symp. **3**: 211.

29. **Meyerhof, O., & W. Möhle**
1936. Biochem. Z. **284**: 1.

30. **Dubuisson, M., & W. Schulz**
1938. Pflüg. Arch. **239**: 776.

31. **Lipmann, F., & O. Meyerhof**
1930. Biochem. Z. **227**: 84.

32. **Dubuisson, M.**
1938. J. Physiol. **94**: 461.

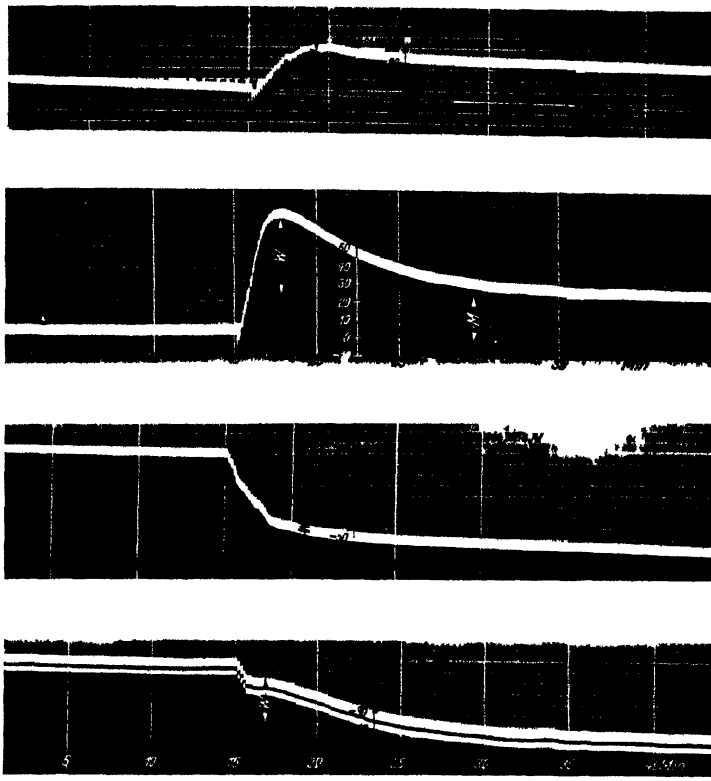
33. **Dubuisson, M.**
1940. Arch. Int. Physiol. **50**: 203.

34. **Margaria, P.**
1934. J. Physiol. **82**: 496.
35. **Engelhardt, W. A., & M. N. Ljubimova**
1939. Nature **144**: 668.
36. **Kalckar, H. M.**
1944. J. Biol. Chem. **153**: 355.
37. **Polis, B. D., & O. Meyerhof**
1946. J. Biol. Chem. **163**: 339.
38. **Price, W. H., & C. F. Cori**
1946. J. Biol. Chem. **162**: 393.
39. **Cori, C. F.**
1946. J. Biol. Chem. **165**: 395.
40. **Hill, A. V.**
1938. Proc. Roy. Soc. London B **126**: 136.

PLATE 7

PLATE 7

Volume change during fatigue of normal muscles (*a* and *b*), and of muscles poisoned with iodoacetic acid (*c* and *d*) *a* and *c* Ringer solution, *b* and *d* paraffin oil



MEYERHOF MAIN CHEMICAL PHASES OF MUSCLE RECOVERY

CHEMICAL PROCESSES OF OXIDATIVE RECOVERY

By Severo Ochoa

New York University College of Medicine, New York, N. Y.

The breakdown of adenosinetriphosphate (ATP) is the energy-yielding process most directly connected with muscular contraction. Since the ATP store in muscle is relatively small, the main function of the recovery processes, both anaerobic and aerobic, is to regenerate this store. As regards the ultimate objective, there is no fundamental difference between anaerobic and aerobic metabolism. But there is an essential quantitative difference, namely, that aerobic metabolism is much more economical, so that more work can be carried out for a given consumption of fuel aerobically than anaerobically.

There is also a difference concerning our knowledge of the anaerobic and aerobic mechanisms. Whereas, in the former case, a rather complete picture can be drawn of the chemical processes and of the manner in which they are coupled with resynthesis of ATP, we still lack a good deal of information regarding the latter.

There is still some degree of confusion as to whether the oxidative breakdown of carbohydrate in muscle is totally different from the anaerobic, or whether, in all cases, muscle activity involves formation of lactic acid and oxidation starts with lactic acid. There is now little doubt that the breakdown of carbohydrate by phosphorylation and cleavage to the triose level occurs in the same manner both in anaerobiosis and aerobiosis. The difference lies in the mechanism by which the phosphorylated triose is oxidized to the pyruvic acid level, and in the subsequent fate of pyruvic acid. Aerobically as well as anaerobically, the phosphorylated triose is oxidized by the corresponding pyridine nucleotide dehydrogenase, and diphosphopyridine nucleotide (DPN) is reduced. In the absence of oxygen, hydrogen (and electrons) from reduced DPN is transferred to another molecule of phosphorylated triose, which is thereby reduced to phosphoglycerol. This reaction is catalyzed by the glycerophosphate dehydrogenase. As soon as pyruvic acid is formed, the reduced DPN is predominantly reoxidized by transfer of its hydrogen to pyruvic acid in a reaction catalyzed by lactic dehydrogenase, so that pyruvic acid is reduced to lactic acid. This is understandable, because the pyruvic lactic system has a higher oxidation-reduction potential than the triosephosphate-glycero-

phosphate system, and the potential of these two systems, in turn, is higher than that of the phosphotriose-phosphoglyceric acid system. However, in the presence of oxygen, hydrogen from dihydroDPN is transported to oxygen over the flavoprotein and cytochrome systems whose potentials are much higher than those of the lactic-pyruvic system. These facts are represented diagrammatically in FIGURE 1.

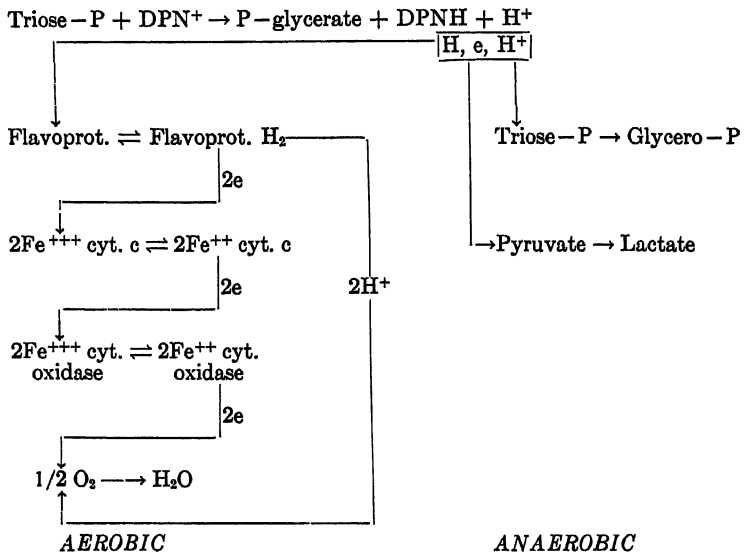


FIGURE 1. Aerobic and anaerobic mechanisms of hydrogen and electron transfer in the oxidation of triose to the glyceric acid level in muscle. (Although not represented in this diagram, oxidation of triosephosphate—glyceraldehyde-3-phosphate—involves uptake of one molecule of inorganic phosphate and leads to 1,3-diphosphoglyceric acid).

Aerobically, pyruvic acid is no longer reduced to lactic acid but is itself oxidized to CO_2 and H_2O. When lactic acid has accumulated as a result of anaerobic activity, it is then oxidized aerobically by lactic dehydrogenase over the flavoprotein-cytochrome system to pyruvic acid.

There is some evidence that muscle respiration may take routes different from the one here sketched for carbohydrate. This evidence is mainly based on experiments with tissues poisoned with iodoacetic acid, which, as is well known, inhibits the enzymatic dehydrogenation of the phosphotriose by blocking essential SH groups of the enzyme protein.[1] It has been observed[2] that iodoacetic acid concentrations which are sufficient to block glycolysis do not inhibit completely the respiration of muscle. While such experiments indicate that there may be

pathways of carbohydrate oxidation in muscle other than over hexosediphosphate and phosphotriose, they do not prove that such pathways are essential, for they may be only emergency or secondary mechanisms. Very little or nothing is known of these other pathways. Oxidation over the hexosemonophosphate-phosphohexonic acid system, present in red blood cells and yeast,[3-6] has not yet been shown to occur in muscle, and the same can be said of the oxidation of sugar to hexonic acid, catalyzed by an enzyme present in liver.[7] Muscle, however, is known to be able to oxidize ketone-bodies, and recent work indicates that fatty acid and carbohydrate oxidation merge into a common pathway in muscle and other tissues.

REGENERATION OF ADENOSINETRIPHOSPHATE

Dr. Meyerhof has already sketched the mechanisms by which glycolysis leads to regeneration of ATP.[8] The two steps concerned are the oxidation of phosphotriose, with uptake of inorganic phosphate, to 1, 3-diphosphoglyceric acid (which reacts with ADP* to form ATP and 3-phosphoglyceric acid), and the reaction of phospho(enol) pyruvic acid with ADP to form ATP and pyruvic acid. In this way, four energy-rich phosphate bonds are generated for every molecule of hexosediphosphate glycolyzed. Starting with glycogen, one phosphate is introduced by the phosphorylase reaction, but another phosphate (the one converting fructose-6-phosphate to fructose-1,6-diphosphate) is introduced at the expense of one energy-rich phosphate bond from ATP. The net result is that cleavage of one glycogen equivalent to two lactic acids generates three energy-rich phosphate bonds. This is what happens in muscle extracts, but, in intact muscle, each glycogen equivalent glycolyzed to two lactic acids gives rise to a net generation of four phosphate bonds.[8] The mechanism by which this is achieved is still unknown. The free energy change ($\triangle F$) of the splitting of one glycogen equivalent to two lactic acids is about — 50,000 gram-calories, and, since the free energy of the pyrophosphate bond in ATP is of the order of 12,000 calories, the formation of four such bonds corresponds to + 48,000 calories. In other words, most of the energy of the anaerobic splitting is utilized for resynthesis of ATP by intact muscle.

The coupling between respiration and formation of phosphate bonds was first observed by Engelhardt,[9] in experiments with blood hemolysates; subsequently, by Lennerstrand and Runnström,[10] with yeast juice; by Kalckar,[11] with kidney extracts; by Belitzer,[12] and by Cori

* Adenosinediphosphate.

and his associates,[13, 14] with heart muscle extracts; and by Ochoa,[15] with brain and heart extracts. In all these experiments, consumption of oxygen is linked with the uptake of inorganic phosphate by phosphate acceptors such as glucose, fructose-6-phosphate, and creatine. There are two steps involved: (a) phosphorylation of adenylic acid or ADP to ATP coupled with oxidation; and (b) transfer of phosphate from ATP to the phosphate acceptor, independent of oxidation.[16] Lipmann[16] observed generation of ATP from adenylic acid coupled with dehydrogenation of pyruvic acid in bacterial preparations.

Our own experiments, using heart muscle extracts, indicate that the complete oxidation of pyruvic acid to CO_2 and H_2O is coupled with the formation of six energy-rich phosphate bonds per molecule of oxygen consumed.[14] Thus, the reaction, Pyruvate + 2.5 O_2 = 3 CO_2 + 2 H_2O, can generate up to 15 phosphate bonds. It is possible that the generation of phosphate bonds by glucose oxidation proceeds with the same efficiency, so that the overall reaction, Glucose + 6 O_2 = 6 CO_2 + 6 H_2O, could generate up to 36 phosphate bonds equivalent to about 430,000 calories. Since the free energy change of the above reaction is about 688,000 calories,[17] this would correspond to about 60% conversion into phosphate bond energy. It is interesting to observe that the oxidation-reduction potential span between the pyruvate-acetate system and cytochrome *c*, of about 0.9 volts, is about two-thirds of that between pyruvate-acetate and oxygen[10]: *i.e.*, some 65 per cent of the energy released by the transfer of one pair of metabolic hydrogens to oxygen is covered by the steps between metabolite and cytochrome *c*. It might be that, between these two levels, the energy released by the process can be converted into phosphate bond energy with nearly 100 per cent efficiency, and that the remaining 35 per cent is evolved as heat.

There is, thus, some evidence that, in terms of phosphate bond generation, the efficiency of the aerobic cleavage of carbohydrate in muscle is about 60 per cent of the anaerobic efficiency. Since the aerobic free energy change (— 688,000 calories per mole) is about 14 times higher than the anaerobic (— 50,000 calories per mole), muscle could theoretically perform, aerobically, a given amount of work consuming only about one-eighth of the carbohydrate that would be required to perform the same amount of work under anaerobic conditions.

It is well known that, under anaerobic conditions, the rate of consumption of carbohydrate by cells is six to eight times faster than

under aerobic conditions. This decreased rate of utilization in the latter case is known as the *Pasteur effect.* The fundamental observations of Meyerhof established the fact that, when lactic or pyruvic acid is oxidized in muscle, only one-fourth to one-fifth of the total amount disappearing is oxidized to CO_2 and H_2O, the remainder being converted to carbohydrate. The mechanism of the Pasteur effect was thought by Meyerhof to consist of a continuous aerobic resynthesis of carbohydrate from its anaerobic split products. More recently, evidence has been obtained suggesting that this cannot be the exclusive mechanism, but that there must be an alternative way.[18] Such a way is visualized as a slowing on inhibition, under aerobic conditions, of the rate of some of the initial steps of the biological breakdown of carbohydrate.

The resynthesis of lactic acid to carbohydrate, as we now know, involves its oxidation to pyruvic acid and the phosphorylation of the latter, by ATP, to phospho(enol)pyruvic acid,[19] followed by conversion to 3-phosphoglyceric acid, phosphorylation of the latter to 1,3-diphosphoglyceric acid, and reduction of this compound to phosphotriose. Thus, two energy-rich phosphate bonds are needed for the synthesis of one glucose equivalent (phosphotriose) from pyruvic acid. Since oxidation of one mole of pyruvic acid may generate 15 phosphate bonds, it can bring about the resynthesis of 6 to 7 further molecules of pyruvic acid to carbohydrate; *i.e.*, out of 6 or 7 molecules of pyruvate disappearing, one would be completely oxidized to CO_2 and H_2O, while the remainder would be converted to carbohydrate. This would correspond to an oxidation quotient of 6 to 7, not far from Meyerhof's experimental findings.

OXIDATION PATHWAYS AND MECHANISM OF HYDROGEN AND ELECTRON TRANSPORT

There is increasing evidence to indicate that the biological oxidation of pyruvic acid and of fatty acids, and the biological cleavage of acetoacetic acid, yield acetyl groups which are further metabolized, through condensation with oxaloacetic acid, to form a six-carbon tricarboxylic acid.[20] The tricarboxylic acid is then oxidized stepwise to oxaloacetic acid, which re-enters the cycle. This is represented in FIGURE 2.

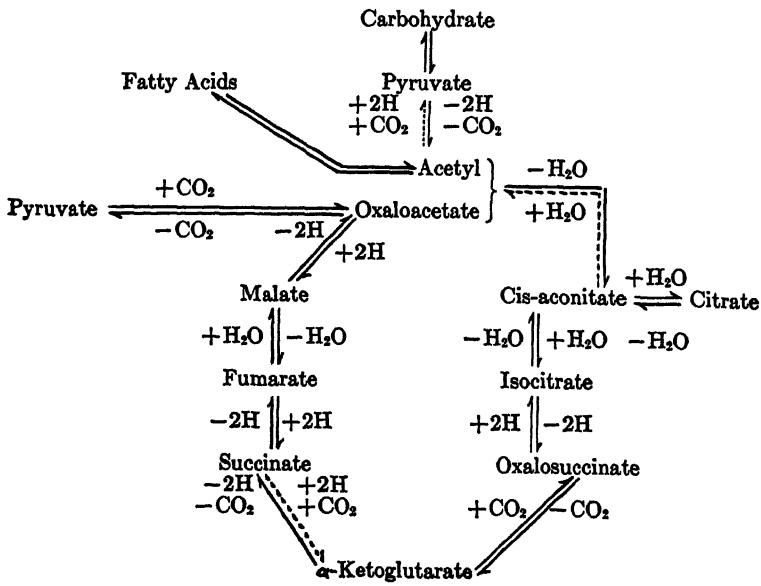
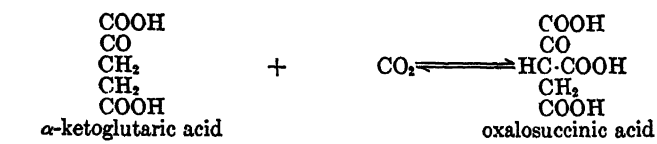
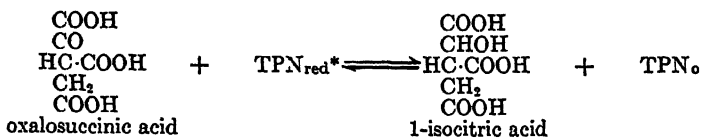


FIGURE 2. Tricarboxylic cycle. (Progress of reactions in the directions indicated by dotted arrows may occur, but has not been demonstrated).

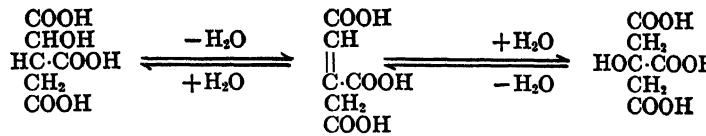
Many of the intermediate reactions of the tricarboxylic cycle are reversible. Of special interest is the reversibility of some of the decarboxylations involved, since the equilibrium of such reactions is very far in the direction of decarboxylation. However, they can be easily reversed when coupled with suitable energy-yielding reactions. As an example, the enzymic synthesis of citric acid from α-ketoglutaric acid and carbon dioxide may be mentioned. The individual reactions and enzymes involved are the following:[21, 22]



Oxalosuccinic Carboxylase



Isocitric Dehydrogenase



and the overall reaction is:

$$\alpha\text{-ketoglutaric acid} + CO_2 + TPN_{red} \rightleftharpoons \text{citric acid} + TPN_{ox}.$$

When this system is combined with the hexose monophosphate dehydrogenase system of Warburg and Christian:[3]

$$\text{Glucose-6-phosphate} + TPN_{ox} \rightleftharpoons \text{6-phosphogluconic acid} + TPN_{red},$$

the following coenzyme-linked dismutation is obtained:

$$\alpha\text{-ketoglutaric acid} + CO_2 + \text{glucose-6-phosphate} \rightleftharpoons \text{citric acid} + \text{6-phosphogluconic acid.}$$

The kinetic and thermodynamic aspects of such reactions have been discussed in a recent article.[23] The course of the above dismutation can be followed manometrically in Warburg vessels, owing to the liberation of CO_2 from bicarbonate by the carboxyl group of phosphogluconic acid, since the CO_2 fixed by ketoglutaric acid is balanced by the liberation of CO_2 from bicarbonate by the third carboxyl group of the citric acid formed.[24] Whether these reversibilities play any role in the metabolism and function of muscle and other animal tissues is still unknown, but they are undoubtedly important in the photosynthetic and chemosynthetic mechanisms of plants and bacteria.[23]

In the case of metabolites that are dehydrogenated by pyridine nucleotide dehydrogenases, the transport of hydrogen (and electrons) to molecular oxygen involves a flavoprotein, cytochrome *c*, and cytochrome oxidase. Although this can be stated with certainty only for TPN systems, such as the isocitric dehydrogenase,[25] there is a factor in crude muscle preparations that catalyzes the reoxidation of reduced DPN by cytochrome *c*,[26] and this factor is in all probability a flavoprotein. Such a factor is also present in yeast.[27] Succinic dehydrogenase and ketoacid dehydrogenases, such as α-ketoglutaric dehydrogenase,[28] react with cytochrome *c*, but other links in the chain are still largely unknown. A scheme of hydrogen (and electron) transport to molecular oxygen in carbohydrate oxidation is presented in FIGURE 3.

* Reduced triphosphopyridine nucleotide.
** Oxidized triphosphopyridine nucleotide.

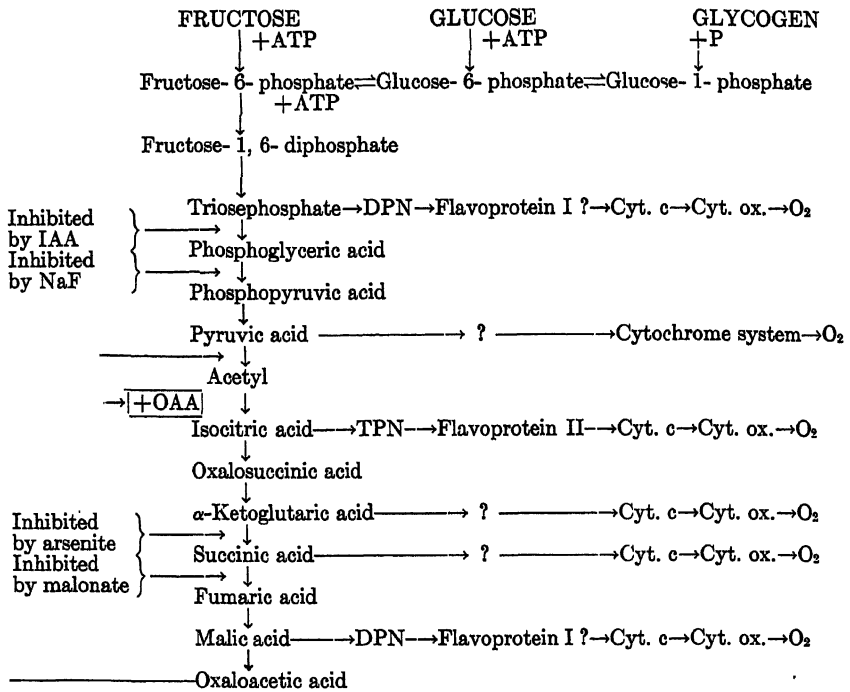


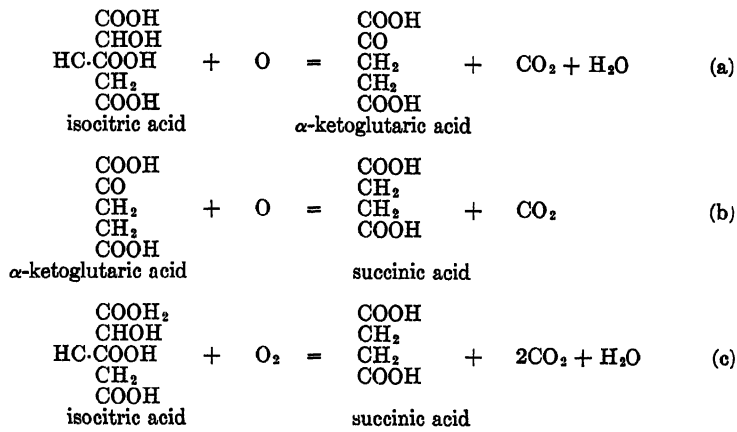
FIGURE 3. Intermediate catalysts involved in the transfer of hydrogen and electrons to molecular oxygen in the stepwise oxidation of carbohydrate in muscle.
Abbreviations: IAA = iodoacetic acid; OAA = oxaloacetic acid; Cyt.c = cytochrome *c*; Cyt.ox = cytochrome oxidase.
The factor of **Lockhart & Potter,**[26] and **Altschul *et al*.**[27] is referred to as Flavoprotein I with a question mark. This factor is specific for DPN. Flavoprotein II, specific for TPN, is the cytochrome reductase of **Haas, Horecker, & Hogness** (cf.[25]). Whether cytochromes *a* and *b* are involved in any of the steps is not definitely known. The important discovery has recently been made that the phosphorylation of glucose by ATP to glucose-6-phosphate, catalyzed by the enzyme hexokinase, is under the dual control of hormones from the anterior pituitary and adrenal cortex on one side and insulin on the other.[30, 31]

INTERMEDIATE OXIDATIVE REACTIONS AND PHOSPHATE-BOND GENERATION

If we consider the oxidative regeneration of ATP in terms of the intermediate reactions of the tricarboxylic cycle, we find that three phosphate bonds are generated in the oxidation of α-ketoglutaric acid to succinic acid and CO_2,[28] but the oxidation of isocitric to oxalosuccinic acid does not generate any,[29] and the oxidation of succinic to fumaric acid probably generates only one phosphate bond.[15] Thus, only a fraction of the total phosphate-bond generation obtained when pyruvic acid is completely oxidized to CO_2 and H_2O is accounted for (cf. p. 838).

The lack of coupling of isocitric acid oxidation with phosphorylation was established with pig heart extracts[29] prepared as in earlier work.[28] In the presence of malonate, to inhibit the succinic dehydro-

genase, isocitrate is oxidized to succinate + 2 CO_2, with uptake of one molecule of oxygen. This is the combined result of reactions (a) and (b): *i.e.*, reaction (c).



In this system (reaction c), glucose is phosphorylated to hexosediphosphate with uptake of inorganic phosphate, but this is exclusively due to coupling of reaction (b) with phosphorylation, and not at all to a coupling of reaction (a), as is clearly shown by the experiments of TABLE 1. When reaction (b) is inhibited by arsenite, there is no phos-

TABLE 1

LACK OF COUPLING OF ISOCITRATE OXIDATION WITH PHOSPHORYLATION

Samples containing pig heart extract and various additions were incubated for 35 minutes in oxygen at 36.5°. All contained 0.055 M phosphate buffer pH 7.0, 0.025 M glucose as phosphate acceptor, and 0.018 M NaF, along with catalytic amounts of $MgCl_2$, $MnCl_2$, and ATP. Concentration of substrates: Citrate, 0.024 M; α-ketoglutarate, 0.016 M. Concentration of inhibitors: Malonate, 0.045 M; arsenite, 0.009 M. Citrate was used instead of isocitrate, because the extracts were rich in aconitase. All values are expressed in micromoles.

Reaction	Substrate	Inhibitors	O_2 uptake	Citrate utilized	α-ketoglutarate		P esterified	Ratio P esterified to ketoglutarate utilized
					formed	utilized		
(a)	Citrate	Malonate Arsenite	12.0	24.3	24.6	0.0	0.0	—
	α-ketoglutarate	Malonate Arsenite	1.9	—	—	4.0	0.0	—
(b)	α-ketoglutarate	Malonate	19.2	—	—	29.6	16.4	0.57
(c)	Citrate	Malonate	26.3	31.0	10.9	20.1*	12.0	0.60

* Difference between citrate utilized and α-ketoglutarate formed.

phorylation (TABLE 1). As explained elsewhere,[15, 28] the low phosphorylation obtained by direct measurement is due to dephosphorylation caused by adenosinetriphosphatase and perhaps other phosphatases, an interference that is higher in pig heart extracts than in the extracts of cat heart used in earlier work.

The mechanism of the aerobic generation of phosphate bonds is obscure. From the work of Warburg on the oxidation of phosphotriose, and that of Lipmann on the oxidation of pyruvic acid by bacteria, we know that, through the formation of an acyl-phosphate bond (1,3-diphosphoglyceric acid, acetyl-phosphate), one molecule of ATP can be formed from ADP for each pair of hydrogens removed from the substrate,[8, 11, 16] but it is difficult to see how the aerobic oxidation of α-ketoglutaric acid to succinic acid and CO_2, when a pair of hydrogens from ketoglutarate react with molecular oxygen to form water, can generate not one but three phosphate bonds.* This difficulty concerns only mechanism, but not thermodynamic possibility, since the release of energy in the passage of two hydrogens from the metabolite, at the bottom of the oxidation-reduction scale, to oxygen, at the top, is large enough to cover the formation of three energy-rich phosphate bonds. This, as discussed above, corresponds to about 60 per cent of the energy release.

BIBLIOGRAPHY

1. **Rapkine, L.**
 1933. C. R. Soc. Biol. **112**: 790, 1294.
 1938. Biochem. J. **32**: 1729.
2. **Barker, S. B., E. Shorr, & M. Malam**
 1939. J. Biol. Chem. **129**: 33.
3. **Warburg, O., W. Christian, & A. Griese**
 1935. Biochem. Z. **282**: 157.
4. **Warburg, O., & W. Christian**
 1936. Biochem. Z. **287**: 440.
5. **Lipmann, F.**
 1936. Nature **138**: 588.
6. **Dickens, F.**
 1936. Nature **138**: 1057.
 1938. Biochem. J. **32**: 1626.
7. **Harrison, D. C.**
 1931. Biochem. J. **25**: 1016.
 1933. Biochem. J. **27**: 382.
8. **Meyerhof, O.**
 1944. Ann. N. Y. Acad. Sci. **45**(9): 377.
 1947. Ann. N. Y. Acad. Sci. **47**(6): 815.
9. **Engelhardt, W. A.**
 1930. Biochem. Z. **227**: 16.
 1932. Biochem. Z. **251**: 343.

* Cf. [15] (especially 1943, p. 503) for a more detailed discussion.

10. **Lennerstrand, A., & J. Runnström**
1935. Biochem. Z. **283**: 12.
11. **Kalckar, H. M.**
1939. Biochem. J. **33**: 631.
1941. Chemical Reviews **28**: 71.
1944. Ann. N. Y. Acad. Sci. **45**(9): 395.
12. **Belitzer, V. A., & E. T. Tsibakowa**
1939. Biokhimiya **4**: 516.
13. **Colowick, S. P., M. S. Welch, & C. F. Cori**
1940. J. Biol. Chem. **133**: 359.
14. **Colowick, S. P., H. M. Kalckar, & C. F. Cori**
1941. J. Biol. Chem. **137**: 343.
15. **Ochoa, S.**
1941. J. Biol. Chem. **138**: 751.
1943. J. Biol. Chem. **151**: 493.
16. **Lipmann, F.**
1939. Cold Spring Harbor Symposia **7**: 248.
1941. Adv. Enzymology **1**: 99.
1944. J. Biol. Chem. **155**: 55.
17. **Ball, E. G.**
1944. Ann. N. Y. Acad. Sci. **45**(9): 363.
18. **Lipmann, F.**
1942. Pasteur effect. In: A Symposium on Respiratory Enzymes: 48. University of Wisconsin Press.
19. **Lardy, H. A., & J. A. Ziegler**
1945. J. Biol. Chem. **159**: 343.
20. **Lardy, H. A., & C. A. Elvehjem**
1945. Ann. Rev. Biochem. **14**: 1.
21. **Ochoa, S.**
1945. J. Biol. Chem. **159**: 243.
22. **Ochoa, S. & E. Weisz-Tabori**
1945. J. Biol. Chem. **159**: 245.
23. **Ochoa, S.**
1946. Enzymic mechanisms of carbon dioxide assimilation. In: **D. E. Green,** Currents in Biochemical Research: 165. Interscience Publishers. New York.
24. **Ochoa, S.**
1946. Fed. Proc. **5**: 148.
25. **Ochoa, S.**
1945. J. Biol. Chem. **160**: 373.
26. **Lockhart, E. E., & V. R. Potter**
1941. J. Biol. Chem. **137**: 1.
27. **Altschul, A. M., H. Persky, & T. R. Hogness**
1941. Science **94**: 349.
28. **Ochoa, S.**
1944. J. Biol. Chem. **155**: 87.
29. **Ochoa, S., & E. Weisz-Tabori**
1945. Unpublished experiments.
30. **Price, W. H., C. F. Cori, & S. P. Colowick**
1945. J. Biol. Chem. **160**: 633.
31. **Price, W. H., M. W. Slein, S. P. Colowick, & G. T. Cori**
1946. Fed. Proc. **5**: 150.

MUSCULAR CONTRACTION

PART IV

MECHANO-CHEMICAL COUPLING

INTRACELLULAR INORGANIC IONS AND MUSCLE ACTION

By H. Burr Steinbach

Department of Zoology, Washington University, St. Louis, Missouri

Inorganic ions, particularly cations, have long been regarded as potent agents for altering and controlling the activities of living cells. The pharmacological literature is replete with studies on effects of altering environmental concentrations of single ions and of ion ratios. On the basis of such studies, a general feeling has grown up that the ions are important controlling factors, a feeling that is well justified, so far as what may be called surface effects are concerned. However, to deduce from such experiments the roles of intracellular cations may not be justified, and it is with intracellular cations that we are concerned. There is, indeed, very little direct information about the actions of cations within the cells. For example, potassium is well known for its effects on irritability and as an agent causing contracture, when supplied in excess in the medium surrounding muscle fibers. However, these effects are usually noted within very short times after the application is made and certainly are not due to any considerable changes in potassium within the fibers themselves. Unless one postulates that all effects, even on the contractile mechanism, are transmitted from the surface inward, it is necessary to conclude that such studies do not relate to actions of intracellular cations, but solely to the effects of the extracellular cations, either directly, acting on the surface, or indirectly, in upsetting a balance between two phases.

CATION DISTRIBUTION IN MUSCLE TISSUE

In order to evaluate the role of intracellular cations in muscle action, it is necessary to know their distribution within the tissue and the magnitudes of the changes that can be induced by various experimental modifications.

In TABLE 1 are assembled representative analyses of muscles from different animal groups, an attempt being made to select invertebrate muscles which compare roughly in function with vertebrate skeletal muscles. For example, the retractor muscles of *Thyone* and *Phascolosoma* are rapidly acting muscles arranged in bands or bundles similar

to vertebrate striated muscles, rather than as sheets of tissue resembling vertebrate smooth muscle. While most of such muscles would histologically be classified as smooth, the adductor muscle of *Pecten* is not only rapidly acting, but is striated as well. For comparison, vertebrate heart and smooth muscle are included in the table.

Inspection of TABLE 1 brings to light interesting differences between

TABLE 1

CATION COMPOSITION OF MUSCLE

Concentrations are given as millimols per kilogram wet weight. The figures in the "ratio" column were calculated by dividing the tissue concentrations by the corresponding ion concentrations in amphibian blood plasma, mammalian blood plasma, or sea water. Values for blood plasma were taken from Conway (1945), assuming rat plasma to be representative for mammals. Chloride contents are included in the table for comparison.

	K		Na		Ca		Mg		Cl	
	Conc.	Ratio	Conc.	Ratio	Conc.	Ratio	Conc.	Ratio	Conc.	Ratio
Vertebrate—Striated										
Frog[1]	83.0	33	25.4	0.24	2.5	1.2	8.3	6.9	10.9	0.15
Rat[2]	101.4	16	26.6	0.18	1.5	0.48	11.0	6.9	16.3	0.14
Cat[3]	113.5	18	21.4	0.15	0.87	0.28	11.6	7.2	13.5	0.12
Beef[4]	84.4	14	18.3	0.13	0.67	0.22	7.8	4.9		
Human[5]	63.0	10	44.5	0.30	1.8	0.58	8.0	5.0	31.1	0.27
Vertebrate—Heart										
Human[5]	70.9	11	37.6	0.26	1.6	0.52	8.7	5.4	33.8	0.28
Dog[6]	77.1	12	32.8	0.22					30.4	0.26
Beef[4]	72.4	12	34.4	0.24	0.95	0.30	6.0	3.8		
Vertebrate—Smooth										
Beef uterus[4]	29.1	5	95.3	0.65	2.67	0.85	3.7	2.3		
Beef bladder[4]	77.7	13	58.2	0.40	1.27	0.41	4.9	3.0		
Invertebrate										
Phascolosoma[7]	106.0	11	122.0	0.28	8.5	0.95			91.0	0.17
Pecten[8]	107.0	11	184.0	0.42	6.5	0.72	26.0	0.49		
Cardium[8]	83.0	8	210.0	0.48	17.0	1.9	35.0	0.66		
Caudina[9]	138.0	14	191.0	0.44	17.0	1.9	8.0	0.15	122.0	0.24
Thyone[10]	169.0	17	173.0	0.39	14.0	1.6			204.0	0.40

[1] **Fenn, W. O.** Physiol. Rev. **16**: 450. 1936.
[2] **Conway, E. J.** Biol. Rev. **20**: 56. 1945.
[3] **Fenn, W. O., *et al.*** Am. J. Physiol. **121**: 595. 1938.
[4] **Wilkins, W.** Proc. Soc. Exp. Biol. & Med. **31**: 1117. 1933–34.
[5] **Cullen *et al.*** J. Biol. Chem. **102**: 415. 1933.
[6] **Muntwyler, E., *et al.*** J. Biol. Chem. **134**: 367. 1940.
[7] **Steinbach, H. B.** Biol. Bull. **77**: 444. 1940.
[8] **McCance, R. A., & H. L. Shipp.** J. Marine Biol. Assoc. **19**: 293. 1933–34.
[9] From **Heilbrunn**, Outline of General Physiology.
[10] **Steinbach, H. B.** J. Cell. & Comp. Physiol. **15**: 1. 1940.

the ions. Comparing vertebrate striated and invertebrate muscles, sodium ion concentrations vary greatly, while the ratios of external concentrations to internal show less variation and are less than 0.5. Chloride concentrations show similar variations, while the ratios are, if anything, even less variable than for sodium, the highest being 0.4. Many detailed studies have shown, of course, that this is because the greater fraction of NaCl is in the extracellular fluid and the variation in tissue content of either ion is primarily a reflection of a simple diffusion equilibrium between the environmental fluids and the tissue spaces. The relatively constant ratios, particularly for chloride, are indications that the relative amounts of extracellular spaces are not greatly different for the different muscle types.

In contrast to the behavior of sodium, potassium is concentrated in most muscles ten to thirty times over the environment. Of greater interest is the fact that the absolute concentration of potassium tends to remain relatively constant, regardless of the osmotic strength of the environment. The osmotic concentration of sea water is over four times that of mammalian blood plasma, while there is less than a two-fold difference between the potassium concentrations. From this, one could conclude that potassium within the cells, while it shows some dependence on potassium in the medium, is relatively independent of the *total* external ionic concentration. Potassium is an intracellular ion, but it is not necessarily osmoregulatory.

Calcium and magnesium present special problems. Tissue calcium fluctuates from muscle type to muscle type, both with respect to absolute concentration and to internal-external ratio. Calcium is never greatly concentrated with respect to the environment, and the tissue concentration may be distinctly low. While the higher external concentration of calcium in sea water ($3 \times$ mammalian plasma Ca) is accompanied by higher tissue calcium of the invertebrate muscles, the internal-external equilibrium is not one of simple diffusion since, with tissues from both major types of animals, ion ratios vary from less than one to nearly two. Magnesium, on the other hand, is more closely regulated with respect to absolute concentration. Sea water has over twenty times the magnesium concentration of vertebrate blood, while, for the major types of muscle, the tissue concentration varies only about four-fold. The result of this is that vertebrate muscle contains magnesium in considerably greater concentration than the external medium, while with marine muscle the situation is reversed.

The figures given in TABLE 1 have a direct bearing on questions relating to the roles of the ions in maintaining osmotic pressure. It is

usually stated that the potassium within the muscle fiber is just about sufficient to establish osmotic equilibrium with the external medium, the implication being that this is one of the major functions of the ion. When a frog muscle, for example, is treated with hypertonic Ringer's solution, the fiber shrinks and the intrafibral potassium concentration increases correspondingly. While this indicates an overall osmotic behavior of the fibers, it does not necessarily mean that potassium must normally serve to maintain the osmotic concentration. If potassium serves an osmotic function, then potassium concentration within the cell, no matter what the concentrating mechanism is, should be regulated to show a proportionality between the total osmolar con-

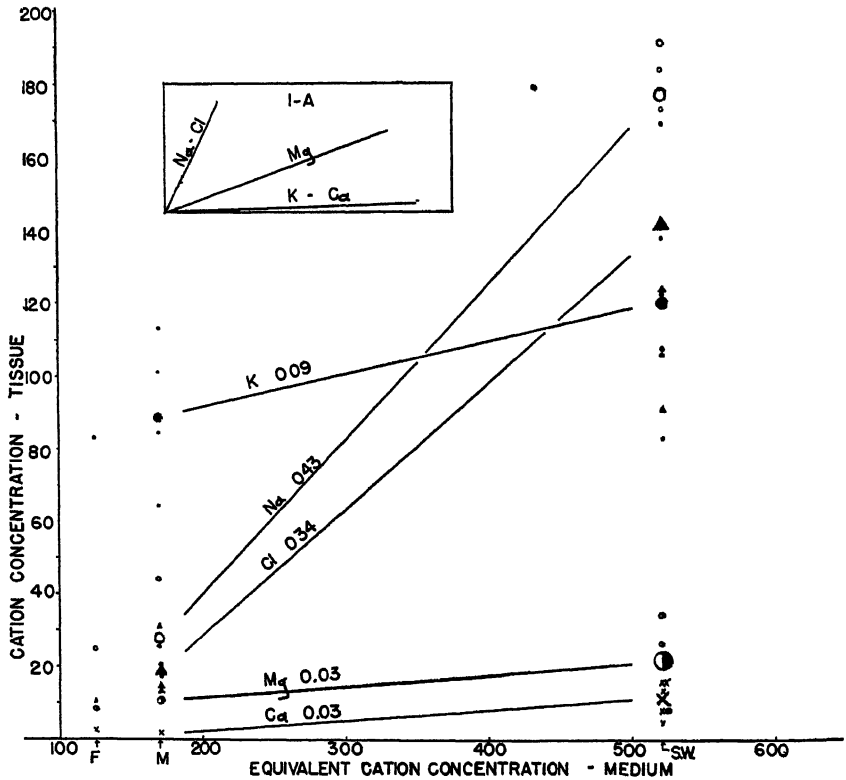


FIGURE 1 Ion concentrations of marine invertebrate, mammalian, and amphibian muscles as functions of the concentration of the media. Tissue concentrations as in TABLE I. Concentration of medium in millequivalents cation per kilogram. For Ca and Mg, the slopes would be slightly greater if tissue concentrations were plotted as millequivalents instead of millimols. Small symbols, individual figures from TABLE I; large symbols, averages. The lines in the small inset give the approximate slopes for the ions of the medium plotted against cation equivalents of the medium.

● = K; ○ = Na; X = Ca; ◑ = Mg; ▲ = Cl; F = frog; M = mammal; S. W. = sea water

centration of the medium and intracellular potassium. It has already been pointed out that comparative figures show this is not the case. Upon comparing different muscle types, it is found that the absolute potassium concentration of the tissue tends to adjust to the potassium concentration of the medium, but not to the total osmolar concentration. FIGURE 1 illustrates this point for potassium, as well as for other ions. Sodium and chloride show proportionate changes, as we go from muscles of vertebrates to those of marine invertebrates. These changes are of a magnitude that would be expected if the ions were in simple diffusion equilibrium with the environment and served to maintain osmotic pressure of some parts of the tissues. On the other hand, potassium, calcium, and magnesium show only slight variations, as the total osmolar concentration of the medium shifts over a wide range.

This situation, as would be expected, leads to an ion deficit within the fibers of invertebrate muscle. No explanation of this is apparent, and it may be suggested that it is not an important point. The osmotic behavior of a cell depends on many factors and, until all the factors can be exhaustively evaluated, apparent discrepancies in calculated osmotic concentrations should not cause undue worry. It seems certain that the "empty sack" implied in most studies of osmotic behavior does not exist in biology, and the whole question of the possible effect of the structural arrangement of protoplasm on volume changes remains to be clarified.

INTRACELLULAR-EXTRACELLULAR ION BALANCE

Calcium

The calcium content of muscle tissue, in general, does not differ greatly from that of the environment. With certain invertebrate muscles (Steinbach, 1940a) and with heart muscle (Krogh *et al.*, 1944), part of the normal calcium content can be washed out, removing around half of the calcium normally assigned to the cells. With the invertebrate muscles, excess calcium in the medium leads to an increase in cellular calcium, with certain limits. Fenn *et al.* (1938) report that stimulation does not appreciably alter the calcium content of cat muscles, while Lissak and Kovacs note a gain in the calcium content of cat and guinea pig muscles with fatigue. Other literature has been reviewed previously (Steinbach, 1940b).

Magnesium

In frog muscle, magnesium is found concentrated several times over that of the environment. Conway and Cruess-Callaghan (1937) and Fenn and Haege (1942) note that immersion of muscles in magnesium-free solutions leads to a loss of about half the metal normally present in the cells. On the other hand, immersion in solutions of high magnesium content leads to proportional gains in magnesium of the cells over a considerable range. No change in magnesium concentracrease in cellular calcium, with certain limits. Fenn *et al.* (1938)

Potassium

In cells in general, potassium tends to be concentrated several times over the environment. Many studies (see references of Fenn, 1940, and Steinbach, 1940b) have shown that muscles will lose potassium to potassium-free media. The maximum loss that can be sustained without impairing tissue function has not been determined. About a third of the cellular potassium can be lost to a K-free medium and then recovered upon immersion of the muscle in potassium-containing solutions (Steinbach, 1940c). Heppel (1939) has shown that feeding rats on potassium-low diets will result in removal of as much as half of the normal potassium content of the cells. Treatment of rats with Doca (desoxycorticosterone acetate) results in loss of cellular potassium (Miller and Darrow, 1941), as does also shock of various kinds (Tabor and Rosenthal, 1945). While there is no doubt about the ability of muscle fibers to lose potassium under different conditions, the usual vertebrate fibers appear to be saturated with this element, so that excess potassium in the medium, within reasonable limits, does not lead to increases in fiber potassium comparable to the losses that can be observed. The great rise reported on feeding potassium to rats (Bruman and Finkelstein, 1936) has not been confirmed (Heppel, 1940), and some equally astonishing concentrations reported upon adrenalectomy must be confirmed. Vertebrate muscle behaves as though there were a limited number of groups capable of binding potassium, and these groups are nearly saturated under normal conditions. This is not true of heart muscle nor of invertebrate muscle (Steinbach, 1940b), where, apparently, the normal tissue is not saturated with potassium. In *Phascolosoma* muscles, new acid groups appear to be formed to take care of excess potassium introduced into the cell (Steinbach, 1940a). Solutions with excess potassium cause considerable swelling of these muscles along with the uptake of potassium, chloride

entering into the whole tissue in such a fashion as to keep the calculated non-chloride space reasonably constant. This leads to the interpretation that potassium is always associated with organic components which occupy space not accessible to chloride. Boyle and Conway (1941) have made extensive investigations of the potassium content of frog muscles immersed in solutions of high potassium chloride concentrations. As Meigs and Atwood (1916) showed many years ago, such muscles gain both potassium and chloride as they swell. Meigs and Atwood further showed that the chloride did not return to normal when the swollen muscles were placed in normal Ringer's solution, even though the muscle weights finally approximated the originals. Boyle and Conway have not reported on recovery of potassium-poisoned muscles.

Fenn and several other workers (cf. Fenn and Cobb, 1936) have shown that fifteen to twenty per cent of the potassium of the muscle may be lost during extended activity. During subsequent rest periods, the potassium content returns to normal. The same general results are obtained in rats with low-K, high-Na muscles (Heppel, 1940). Studies with radioactive potassium (Fenn *et al.*, 1941) have shown that, over long periods of time, all of the potassium of the mammalian body is exchangeable.

Sodium

Sodium is usually thought of as existing predominantly in the body fluids or the external aquatic environment, although variable and sometimes considerable amounts are found in tissue cells, notably of heart muscle (Yannet and Darrow, 1940). Except in the case of insects (cf. Heilbrunn, 1943, p. 450), sodium is the major cation of body fluids. In muscle tissue, a slight excess of sodium over chloride is usually found. Mond and Netter (1932) have interpreted this as showing that some sodium can stick to the outer layers. In view of the large excess of sodium over potassium found in heart muscle (TABLE 1), it seems improbable that a simple mechanical hypothesis can account for the extra sodium; part of the tissue content of the element must be in the cells.

In every case where parallel measurements of potassium and sodium have been under carefully controlled conditions, a reciprocal relationship has been found of such a nature as to indicate that, when potassium leaves the cells, sodium enters. To select recent instances, this has been noted in mammals with low potassium diets (Heppel, 1939; Miller and Darrow, 1940); following shock (Tabor and Rosenthal,

1945); during contraction (Fenn *et al.*, 1938); and upon treatment of heart with such drugs as digitalis (Wood and Moe, 1942). In isolated frog muscle, the exchange has been noted by Fenn and Cobb, (1934) and by Steinbach (1940c). The reverse exchange, potassium for sodium, in the cell when corrective measures are applied, has not been studied in all cases, but has been noted for the isolated frog muscle (Steinbach, 1940c); for rat muscle (Miller and Darrow, 1940) following injection of KCl into low-K animals; and for the rat, cat, and frog following stimulation (Fenn, 1940).

Chloride

Chloride must be considered briefly, since it is probably related to cation balance in the protoplasm. As noted in TABLE 1, there is usually less chloride than sodium in muscle tissue and, therefore, assuming the extracellular spaces to be equilibrated with body fluids of external environment, there must be less chloride than sodium inside the cells. Several studies have shown, however, that extracellular space, measured by other methods (Mg-space, inulin space, etc.), is smaller than the so-called chloride space (cf. Wilde, 1945). In addition to this, methods commonly used for estimating chloride have recently been called into question (Heilbrunn and Hamilton, 1942; Wilde, 1945).

THE ACTION OF INTRACELLULAR CATIONS

The studies reported on ion balance have shown that individual ion concentrations are regulated by special selective mechanisms in such a fashion that normal physiological variation occurs only over narrow limits, and that it is a rather difficult and slow process to change appreciably the cation composition of the interior of the muscle fiber. Keeping these two points in mind, it becomes apparent, on surveying our knowledge, that we have few, if any, clear-cut demonstrations of effects of intracellular cations on the contractile mechanism. This statement is based on the assumption that irritability can be separated from the contractile processes. While there is no doubt that these functions are closely related, it has been demonstrated with heart muscle and with single striated fibers that the conducting mechanism can be stopped, leaving local graded contractions of an extent depending on stimulus strength. Buchthal *et al.* (1943) discuss evidence that the contraction wave of a single fiber has characteristics quite different from the usual excitation wave. Therefore, while some valuable data may be excluded, this discussion will deal only with cases where an

unequivocal change in protoplasmic cations can be assumed to have taken place.

Some of the difficulties may be illustrated by considering potassium, one of the ions easiest to move to and from the muscle fiber. Potassium causes contracture in muscle fibers when applied in excess. This contracture usually occurs at a potassium level three to four times the normal plasma level and at a concentration far higher than is necessary to abolish normal response to electrical stimuli (Hegnauer *et al.*, 1934). However, the contracture can be elicited by test exposures of only thirty seconds to the high potassium solutions. Obviously, altered surface conditions must be involved. The effect could be related to the irritability mechanisms, rather than to the contractile mechanisms themselves, although the maintenance of contracture and associated chemical changes may be conditioned by high intracellular potassium. Studies of excess potassium on striated muscle have the further limitation that the fibers are already more or less saturated with the element. With heart muscle containing exchangeable Na within the fibers in larger quantity, the effect of potassium appears to be mainly to abolish the normal conducting mechanism (Zwikster and Boyd, 1935). Heart muscle treated with salt solution of about six times the normal K concentration will still respond to electrical stimuli, but in a graded, non-conducting fashion.

During the course of experiments dealing with the production of low-K, high-Na rats, observations have been made to show that there is no obvious correlation between the level of potassium and the ability to perform work (Miller and Darrow, 1941), and that the release of potassium from the fibers in exchange for sodium during activity remains nearly normal (Heppel, 1940). From the data reported, it is impossible to tell whether such things as rate of development of tension or rate of relaxation have been affected or not. Low-K muscles appear to fatigue easily and to develop less total tension (Heppel, 1940). In fact, there are several reports in the literature on the antifatiguing effect of injected potassium chloride. This effect is presumably related to the loss of potassium during contraction, this loss being less if the inside-outside gradient is lower due to the injected salts. In addition, there are experiments (Miller and Darrow, 1940) to show that potassium-depleted animals can withstand a stronger dose of injected potassium than can normal animals, thus indicating that the muscles serve, in part, as a relatively indifferent potassium reservoir which can vary within fairly wide limits. That excess potassium does have internal effects on cells, has been shown by Heilbrunn and his

students (cf. Heilbrunn, 1943), but a direct action on the contractile mechanism of muscle has not been demonstrated. What is needed is a careful and precise study of the reversible effects of altered intrafibral potassium on time and tension constants of the simple muscle twitch.

The situation with respect to the other ions is similar. The two most recent studies on magnesium effects (Fenn and Haege, 1942; Ashkenaz, 1938) note that increased magnesium (shown by Fenn to result in increased intracellular Mg) is accompanied by a considerable rise in rheobase, but no comments are made on tension developed or such factors relating to contraction. It has been shown that calcium, injected into single muscle fibers (cf. Heilbrunn, 1943), will cause marked contracture, but the probability is that the local concentration was greater than could be mustered in the normal life of the cell. There is some evidence that calcium is freed from an un-ionized state during activity (Heilbrunn, 1943).

The effect of calcium in increasing the amplitude of the heartbeat may well be a case of direct action of intracellular calcium on the contractile mechanism. According to figures given by Clark (1938, p. 224), a four- to five-fold increase of calcium in the fluid perfusing a frog heart gives a doubling of amplitude with only a minor change in duration of the beat. Similar effects of excess calcium on voluntary muscle have also been reported.

The experiments adduced to show a general lack of effect on the contractile mechanism of altered intracellular potassium would also apply to sodium, since these two ions have been shown to bear a reciprocal relationship to each other. Protoplasmic sodium can be varied within rather wide limits, without effects that can be ascribed directly to the tension-developing mechanisms.

The importance of the protoplasmic cations for muscle contraction, as distinct from the phenomena of stimulation, thus seems to be indicated, mainly indirectly, by two lines of evidence: (1) the special distribution of the cations between fibers and environmental fluid; and (2) the changes noted during activity. The changes in potassium content of muscle during activity have not been seriously questioned. Fenn *et al.* (1941) have also described experiments which indicate that the rate of penetration of radioactive potassium (permeability) is not significantly altered during activity. Thus, the potassium release from the fiber (and uptake during recovery) must relate to changes within the fiber, rather than to changes in ease with which the cations can cross the surface. This might signify appearance of potassium

ions from some reservoir of bound material opened up by activity, as has long been claimed by Ernst. The case for calcium and magnesium changes during activity is not as well established, although evidence has been adduced to favor an appearance of calcium ions during contraction (cf. Heilbrunn, 1943). The failure to observe changes in calcium and magnesium concentrations in stimulated muscles (Fenn *et al.*, 1938) might be related to the low diffusion gradients operating, as compared to the high gradient for potassium.

Evidence that the cations play a role in normal muscle action may also be gained from histochemical studies of the distribution of minerals. In general, however, localization of materials in muscle by histochemical methods has not led to unequivocal results. A notable exception is the study by Caspersson and Thorell (1942) showing ultraviolet absorbing areas in the *I* disks. Macollum (1905), using a cobalti-nitrate reagent applied to living muscle fibers, claimed that potassium was localized in the *A* bands. Macollum used muscles from a variety of animals, including insects which have widely spaced bands. His results were termed artifacts by Gersh (1938), who used frozen-dried sections of frog muscle having very narrow, closely spaced bands. Gersh claimed that Macollum's results could be obtained when a slight amount of water was added to the dried system before potassium precipitation, the implication being that the potassium was differentially concentrated from a normal, evenly dispersed condition during the water treatment. It is difficult to criticize histochemical work, since the personal element is very important in such delicate techniques, but the reasoning used by Gersh seems open to question. Many of the fibers of a block of muscle are contracted or started to contract during quick freezing (Scott and Packer, 1939). Bureau (1934), using fresh fibers from the frog, and confirming Macollum, claims that the slightest contraction causes an even distribution of potassium, the normal condition being a localization of the element in the *A* bands. It seems quite probable, then, that the banding, observed by Gersh when a little water entered his dried preparations, represents a shift back to what may be a normal condition. In any case, it seems safe to adopt the attitude that the localization shown by Macollum, Bureau, and others does bear a definite relationship to the chemical make-up of the fiber and is probably related to the normal distribution. Dubuisson (1942) has recently reviewed the whole question and feels strongly that potassium is localized in the anisotropic regions.

Mineral distribution has been studied by both microincineration methods and the emission electron microscope (Scott and Packer,

1939). The greatest deposits of ash correspond to the *A* bands, although, in frozen-dried material, appreciable quantities are in the *I* disks. This is in contrast to the findings on ordinary fixed material, which show the deposits in the *A* bands, but little or no ash in the *I*. Minerals appear to be free to be washed out in the *I* bands, but not in the *A*. This finding agrees with the results of the ion balance studies, particularly of calcium and magnesium, which show that only a part of the ion content can be washed out of fresh muscle fibers. Furthermore, in view of the fact that calcium may be present in muscle fibers in a concentration lower than the environmental concentration, one is led to suspect that what Scott and Packer designate as calcium and magnesium may be mostly magnesium. Calcium and magnesium are concentrated in the "contraction bands" and, in the resting muscle, in the *A* disks, although the distribution may not be exclusive. Deucher (1941) has studied glycogen distribution, and claims a localization in the *A* disks. Deucher also confirms Macollum in finding the potassium in the *A* disks.

The diagram published by Verzár (1942) contains most of the probable distributions of special components of an ordinary striated muscle. This diagram shows the *A* bands containing glycogen, myosin, potassium, and the alkaline earths, the *I* bands containing the nonmyosin proteins, ATP, and the intrafibral sodium. The evidence, thus summarized, is strongly indicative of a compartmentalization of strategic material within the striated muscle fiber, a condition very likely to be associated with rapid action and recovery. In contrast, the same general studies have shown that, with smooth muscle, a more uniform visible distribution is encountered. An orientation of active components in smooth muscle may still be assumed to exist, as has been indicated for other types of cells (Spiegelman and Steinbach, 1945), but the less rigid segregation might be thought to offer a less effective controlling system for rapid muscle contraction.

It is of some interest that a selective distribution of cations within the fibers would account for a great many observations on cation balance. For purposes of calculation, we can make the following assumptions about the frog sartorius, following the suggestions of Dubuisson.

1. Extracellular space is about 15 per cent of the total tissue volume.

2. Potassium is concentrated in the *A* bands, the concentration being nearly equal (90 per cent) to the total ionic concentration of the medium (240 meq./liter). We also assumed that potassium is combined with a polyvalent ion, or is "bound." All volume changes due to osmotic pressure variations take place in the *A* bands.

3. The sarcolemma is permeable to all inorganic ions. The distribution of ions between the *I* region and extracellular fluid thus follows a simple Donnan equilibrium. Na is assumed to make up 90 per cent of the cation content of the *I* bands.

The curves in FIGURE 2 show the concentration of potassium, sodium, and sodium plus potassium of the whole sartorius muscle equilibrated

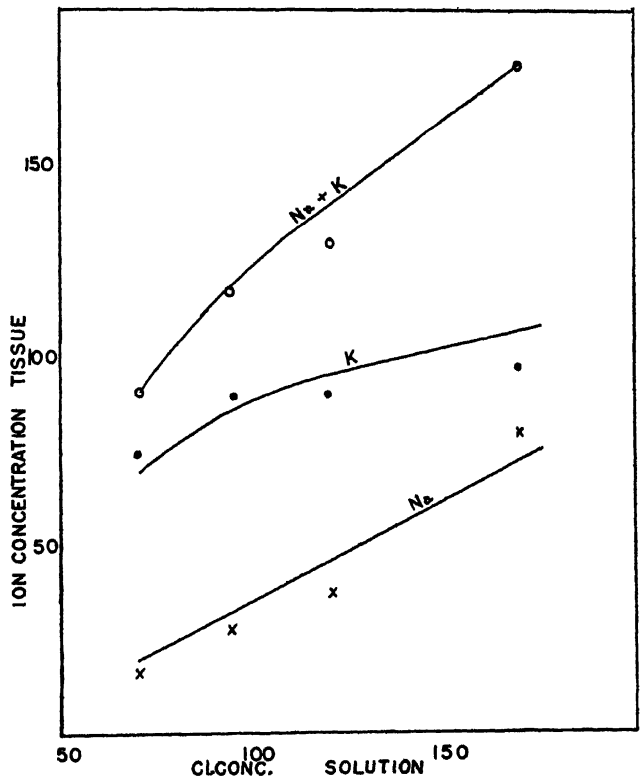


FIGURE 2 Tissue concentrations of sodium (crosses), potassium (dots), and their sums (circles) plotted against concentration of Ringer's fluid in which the muscles were equilibrated several hours (see **Steinbach**, 1944) The curves pass through values calculated on the assumption that the *I* bands of frog muscle fibers are permeable to all ions, while the major part of the osmotic pressure of the *A* bands comes from potassium (see text) Concentrations in millimols per kilogram solution or final weight of tissue.

with Ringer's solution of different NaCl concentrations. The points following the curves represent observed values (Steinbach, 1944). On the whole, the agreement is very good, the deviations at high external Na^+ concentrations being such as would be predicted if some exchange of Na^+ for K^+ of the *A* bands could take place. Chloride values are

not shown, but, assuming a reasonable value for non-diffusible anions of the *I* bands, chloride contents of whole muscle may be calculated with comparable accuracy.

Besides accounting well for the total cation content of muscle, the distribution scheme would explain the otherwise puzzling finding that the rate of exchange of radioactive intracellular sodium of muscle is greater than the rate for potassium. Potassium release upon contraction may be accounted for by assuming that contraction involves a loosening of the structural forces that maintain the normally highly-oriented *A* band condition.

ION EFFECTS ON ENZYME SYSTEMS

There are numerous instances in the literature where the common ions have been shown to influence various steps in the glycolytic systems. Much of this work is summarized by Barron (1943). Boyer *et al.* (1943) show that potassium specifically increases the rate of creatine phosphorylation. This action is in addition to a stimulatory effect of Mg^{++} or Mn^{++}, and is antagonized by Ca^{++}. The optimal effect of K^{+} occurs over a rather wide range of concentrations (from about 10^{-2} to 10^{-1}), but well within the range that could be provided by the normal K^{+} content of muscle, whether part of the potassium was bound or not. If creatine phosphate is concerned with recovery reactions, these findings might well explain the effect of KCl injections on recovery from fatigue. It might also be postulated that a release, during contraction, of potassium ion from a previously bound state facilitates recovery from the contraction. This, in turn, is accompanied by a rebinding of the ion. In an earlier paper, the authors note that a calcium ion concentration of about 10^{-2} inhibits phosphocreatine formation by over 90 per cent. Thus, there is a possibility of a dual control of the recovery reactions in normal muscle. An initial increase of calcium ion (the calcium release cited by Heilbrunn) would inhibit the formation of phosphate donors (concerned with recharging or recovery) during active phases of shortening or development of tension. Accompanying the tension development, potassium ion increases, and calcium rebinding could occur with consequent release or stimulation of the recovery mechanisms. One of the virtues of the work by Boyer *et al.* is that they worked with muscle homogenates, rather than with purified enzyme systems. While this has a disadvantage, in that it is difficult to tell exactly which enzyme is being influenced by an ion, it gives important physiological information in dealing with systems that

more nearly approximate the conditions within the muscle than do mixtures of purified materials In connection with this, it is interesting to note that Boyer *et al.* (1943) find no stimulating effect of Ca on ATP hydrolysis, nor is there any inhibition by oxalate. This behavior has also been reported for the apyrase system from chick embryo material (Moog and Steinbach, 1945).

By far the greatest interest in recent years has centered on myosin and, in particular, on the calcium activation of the ATPase activity (cf. Bailey, 1942). At a very alkaline pH and within a rather narrow range of substrate-enzyme ratios, calcium has a pronounced stimulating effect on the hydrolysis of the terminal phosphate of ATP. Coupled with the finding that myosin is difficult to separate from the enzyme, the discovery of the ATPase activity has led to many investigations, and even more speculations, concerning the role of the protein as a device for the direct conversion of chemical (phosphate-bond) energy to mechanical energy. Engelhardt (1941) considered that the hydrolysis of ATP caused the re-extension (relaxation) of myosin and, hence, of the muscle. Bailey (1942), impressed by the role of calcium, relates the ion effects to the calcium release theory of stimulation, thereby implying that the hydrolysis of the ATP (stimulated by the release of calcium ions) causes the contraction. Bailey, however, does not specifically uphold any one point of view. The question of activity *vs.* recovery methods of coupling has been reviewed by Sandow (1945).

It is probably premature to consider that the normal physiological role of calcium in muscle contraction is related to the activating effect upon the myosin-ATPase system. Banga (1941–42) has shown that all common ions, including both calcium and magnesium, have activating effects on the hydrolysis of ATP The major difference between calcium and magnesium appears to be that they operate over different concentration ranges Banga reports that magnesium activation is optimum at around 10^{-1} M, while calcium acts most strongly at 10^{-2} M Both ions inhibit at high concentrations. In addition, many myosin preparations show virtually no effect of calcium, or even a slight inhibition. Such preparations are not affected by oxalate. In any case, the conditions laid down by Bailey for maximum calcium effect (low enzyme concentration and high pH) could hardly be expected to hold in normal muscle (see, however, Greville and Lehman, 1943).

Additional difficulty arises, particularly with respect to the well-known antagonism of calcium and magnesium. Greville and Lehman (1943) and Banga and Szent-Györgyi have shown that magnesium, 10^{-2}

TABLE 2

Calcium activation and inorganic sodium triphosphate inhibition of the hydrolysis of ATP by myosin. Rabbit myosin preparation, first precipitate washed and redissolved in 0.5 M KCl-$NaHCO_3$, one ml. containing 0.98 mg. N. Hydrolysis tests at 37° C., reaction volumes 1.5 ml. Besides the additions noted in the table, the contents of each tube was 0.03 M in KCl, ca. 0.05 M Barbital buffer, pH 7.3 and excess 7″ ATP phosphorus.

Activity μgP/mgN/min.

Ca concentration	Concentration Sodium Triphosphate			
	0	0.0014	0.007	0.014
0 (0.003M Na oxalate)	56			
0	58	26	20	18
0.0025	65			
0.005	68	59	71	42
0.0075	61			
0.010	61	59	65	59
0.020	60	57	55	57
0.030		65	65	65

M, will inhibit the activation caused by 10^{-2} M $CaCl_2$. This, presumably, relates to the differences in optima noted by Banga; hence, there is no way by which the magnesium inhibition could be overcome by adding more calcium. Excess calcium would merely enhance the inhibition. Greville and Lehman were probably correct in concluding that the usual calcium antagonism to magnesium narcosis relates to effects on the myo-neural junction.

Tests on a large number of rabbit myosin preparations have shown that calcium activation of the ATPase system may be highly variable. In some instances, little or no activation is noted. An example is given in TABLE 2. Not only is the maximum activation less than 20 per cent, but oxalate has little or no effect. The table also includes data on the joint action of calcium and the inhibitor, inorganic sodium triphosphate, on myosin-ATPase. Calcium activation may be simulated perfectly by such systems, and, in all probability, calcium effects on crude extracts may relate to removal of natural inhibitors, rather than to direct activation of an enzyme.

During the course of work on apyrase activity of cytoplasmic granules, calcium activation was noted only under conditions that might be suspected of distorting the granular structure. Similar preliminary tests have been made on muscle. Frog muscle was frozen in dry ice, ground and extracted in the cold with M/10 KCl, pH 7.3. The whole mixture was sent through a homogenizer, and large particles were al-

lowed to settle off, or were centrifuged off at low speed. The resulting cloudy supernatant contained particulate units of varying size and shape, apparently representing fragments of muscle fibers. This particulate material was collected in the cold by centrifuging at about 21,000 g for fifteen minutes, resuspended in M/10 KCl and tested for apyrase activity. The preparation was highly active (TABLE 3),

TABLE 3

Effects of treatment of particulate matter from frozen-ground extracts with 0.5 M KCl. Particulate matter prepared from undialyzed extracts as described in the text. Each reaction tube contained 1.5 ccm. fluid, pH 7.3 (Barbital buffer) 0.07 M KCl with ATP in amounts such that labile phosphate was greater than twice the amount of free phosphate appearing during the test. $CaCl_2$, when added, was 0.003 M. All tests at 37° C. a and b, separate experiments.

	Activity μgP/mgN/min.		$\frac{+Ca}{-Ca} \times 100$
	−Ca	+Ca	
Untreated, a	7.1	5.2	73
Untreated, b	20.4	14.5	71
KCl treated a (3 hrs.)	6.4	8.3	130
KCl treated b (1 hr.)	19.5	23.1	118

but showed an inhibition with calcium. One sample of suspension was then made M/2 in KCl, and left in the cold for three hours, together with a control sample correspondingly diluted with M/10 KCl. Apyrase tests of strong KCl-treated and untreated material, at comparable ionic strengths, showed that the KCl treatment had resulted in a preparation showing calcium activation (TABLE 3).

Additional preliminary observations indicate that the enzymatic activity of the particulate matter depends, perhaps, upon the method of grinding the tissue, as well as upon the extracting medium. Comparable samples of frog muscle were obtained, and one sample was frozen and ground, as described for the experiments reported in TABLE 3. The other sample was ground in the cold, but unfrozen, in a mortar and then in a Waring Blendor. Both samples were homogenized, cleared, and fractionated as described before. Enzymatic activity of both final supernatants and particulate fractions was studied.

The M/10 KCl supernatants both showed fair apyrase activity with marked calcium inhibition. Particulate material showed greater enzymatic activity and less effect of calcium (TABLE 4). With the material ground while still unfrozen, however, the relative activity of the particulate material was much greater, and the activity of the

TABLE 4

Hydrolysis of ATP by muscle extracts prepared in M/10 KCl, pH ca. 7.8, from frozen-ground (A) and unfrozen-ground (B) frog muscle. Extracts cleared of large particles by three-minute centrifuging at 2,000 g. Cleared extracts dialyzed overnight against 10x their volume of H_2O, and then fractionated into particulate matter and supernatant by centrifuging 15 minutes at 21,000 g. Reaction mixtures as in TABLE 3.

	μgP/mgN/minute		$\frac{+Ca}{-Ca} \times 100$
	−Ca	+Ca	
Frozen-ground, soluble	5.9	2.0	34
Frozen-ground, particulate	9.1	6.9	76
Unfrozen-ground, soluble	3.1	0.65	21
Unfrozen-ground, particulate	18.0	16.3	91

soluble portion correspondingly less, than the comparable preparations from frozen-ground material. Since the freezing-grinding technique may be assumed to cause less change in all chemical equilibria of the system, these results would tend to confirm Kalckar's (1942) suggestion that the ATPase is really adsorbed very tightly to myosin, but may be free in the muscle.

The reaction of soluble enzyme systems to ions is of passing interest. At constant ionic strengths in the reaction mixtures, the activity is somewhat greater with potassium than with sodium. The effects are not very striking, however. Calcium and magnesium have been studied (TABLE 5). Calcium shows a strong inhibition at all concentrations tested; magnesium activates at concentrations up to $3.3.10^{-3}$

TABLE 5

Effects of calcium and magnesium of ATP hydrolysis by frozen-ground muscle extracts. Preparation and testing as described in TABLE 4, except that dialysis was carried out against M/10 KCl. Tests on extracts without removal of fine particulate matter.

Concentration of added salt, x 10^{-3}		Activity μgP/mgN/min.
$CaCl_2$	$MgCl_2$	
0	0	3.4
1.6	0	2.7
3 3	0	2.1
6.6	0	2.0
13.2	0	1.7
19.8	0	1.6
0	1.6	5.5
0	3.3	4.7
0	6.6	3.3

M, but higher concentrations inhibit. The soluble enzyme could be best described as an apyrase, since it appears to hydrolyze both labile phosphate groups from ATP. At pH 7.3, activity with phenylphosphate as substrate is less than one-tenth the activity with ATP. In view of these results and those reported by the Szent-Györgi group, it is obvious that the whole phosphatase system of muscle needs much closer examination. The enzymes that can split ATP in muscle are several, and reactions to ions are quite dissimilar for the different components.

Recently, Szent-Györgyi and his group have published a series of papers dealing with the "myosin system" which gives several new reactions to take into account in considering the action of ions in protoplasm. They find that what has been termed myosin consists, in reality, of two water-soluble proteins, myosin and actin, which can reversibly combine to give actomyosin, the common myosin, soluble only in strong salt solution. Actin shows flow birefringence, has pronounced anomalous viscosity, and can exist in an inactive form with properties in solution more nearly resembling myosin. The reactions that exist, according to the Szent-Györgyi group, may be outlined as shown in TABLE 6.

Actomyosin threads (Gerendas, 1941–1942; Szent-Györgyi, 1941–1942), in the presence of traces of salts, will contract vigorously to about one-third the original length, upon the addition of ATP. Relaxation is not spontaneous, but can be induced by altering the ionic environment in such a way as to cause swelling (excess KCl). While the papers are not entirely clear on this point, it appears that the contraction of threads, in the presence of 0.1 M KCl, is less rapid upon the addition of ATP than in the absence of KCl. However, if 10^{-2} M $MgCl_2$ is added to the KCl solution, the ATP contraction is greatly enhanced. Szent-Györgyi (1943) states that calcium will not substitute for magnesium and, in some cases, will inhibit the contraction almost completely.

The work by Szent-Györgyi and his group provides a great deal of information that should stimulate much further interest in the muscle proteins. While the details must await confirmation, it seems apparent that the major outlines are correct, and that the fibrous protein system of muscle is in reality two proteins. One, myosin, is associated with phosphatase activity, and the other, actin, has such a very long molecule that it is probable that it supplies the continuous fiber structure noted by Schmitt and others. ATP, according to the several authors of *The Studies*, combines with myosin; the actomyosin is

TABLE 6

Summarized from "Studies from the Institute of Medical Chemistry, University of Szeged," Vol. 1, 2, 3. Enzyme presumably ATPhase or apyrase. + = solubility or activation, − = precipitation or inhibition. Figures in parentheses give order of magnitude of concentrations (see text).

Thread contraction	Thread formation	Solubility	Components	Enzyme activity
		$+H_2O$ ←	Actin (inactive)	
			←Cations+ Ca, Mg (10^{-2})	
			Na, K (10^{0})	
		H_2O ←	Actin (active)	
		Ca, Mg (10^{-2}) K, Na (10^{0})	+	
		$+H_2O$ ←	Myosin → ?	ATPase
		$+NaK$ (10^{-1})		
		− [Na, K (10^{-2}) Ca, Mg (10^{-3})	ATP	+ [Mg (10^{-1}) Ca (10^{-2}) K-Na (10^{-1})
+ [ATP Mg (10^{-2}) K (10^{-1})		$-H_2O$ ←	Actomyosin (Myosin B)	
		$+K, Na$ (10^{-1})		− [Mg (10^{-3}) Ca (10^{-1}) K-Na (10^{0})
−Ca (10^{-2})				

split; contraction takes place; hydrolysis occurs, and the actomyosin is reformed. Physical effects (thread contraction, change in flow birefringence due to ATP) are noted only with the actomyosin, and not with the single proteins. The order in which the reactions take place is not certain, although Szent-Györgyi appears to favor the idea that the ATP-myosin complex is contractile, possibly without direct transfer of phosphate to the myosin. In precipitation experiments, no phosphate is found bound to myosin, after the addition of ATP. No studies are reported on the possible contraction of actin alone.

The majority of the ion effects described on the actin, myosin, and actomyosin proteins appear to be non-specific, in that sodium will usually substitute for potassium, and magnesium for calcium. With purified myosin, however, as with the impure myosin (actomyosin of

unknown proportions) of earlier work, potassium salts are somewhat better solvents than sodium salts. In general, calcium and magnesium are more potent precipitants of actin and myosin than are potassium and sodium, and the former are also much more effective in activating actin. No great difference is noted between the members of a pair, however. Aside from the different effective concentration ranges of magnesium and calcium in activating the ATPase system, the only really specific ion effects noted are those of magnesium and calcium on thread contraction. In concentration ranges likely to be encountered within the muscle fiber, calcium inhibits thread contraction while accelerating hydrolysis of ATP; magnesium accelerates thread contraction and inhibits ATP breakdown. These findings could be built into a theory of ion action, if we could assume that the ATP binding for contraction involves both a contractile protein and an enzyme, while hydrolysis concerns only the enzyme. If dephosphorylation occurred with excessive speed, the contractile complex might, in a practical sense, never be formed. Magnesium would be effective in promoting contraction of isolated threads by slowing down the rate of hydrolysis, rather than by promoting a reaction directly. This hypothesis would also fit in with the apparent release of calcium ions during contraction. If excess calcium inhibits thread contraction, then the release of calcium from muscle might indicate the removal of excess from the fibers, partial inactivation of the hydrolytic system, and contraction due to the formation of a stable ATP-contractile-protein complex. The obvious major difficulty is that magnesium treatment should actually give contraction, rather than anesthesis. A way around this objection may be offered by the finding (Banga, 1943) that magnesium greatly accelerates other phosphatase systems associated with the muscle proteins, possibly leading to a depletion of local supplies of ATP. Since ATP is presumably localized in the *I* bands of resting muscle, magnesium penetrating freely into these bands could cause such a change.

The evidence that has been summarized so far does not lead to any very far-reaching conclusions about the roles of the intracellular cations. The major points shown are as follows:

1. Intracellular cations are probably distributed (localized) relative to intracellular structures. They are not evenly dispersed. One may infer that, for every equilibrium point between internal and external concentrations of an ion, there is a corresponding structural equilibrium within the cell.

2. Intracellular cations influence a variety of enzyme systems in

contractions likely to be found in muscle. Control of reactions by cations is doubtful, however, unless it can be shown that there is virtual exclusion of a specific cation, as an ion, from a specific region.

3. For those ions adequately tested (potassium and sodium), large changes in intracellular concentrations can be effected without destroying the ability of a muscle to do work.

4. Intracellular cations are in some sort of labile equilibrium with the same cations of the external medium. This equilibrium, for at least some ions, is shifted during activity.

From these points, one can conclude that intracellular cations probably act primarily in relationship to the organic structure of the cell. One may visualize the resting muscle as having the highly oriented protein units of the *A* bands bound tightly together with an associated potassium component. The structure of the *A* bands will then be, in part, a potassium structure, and will be balanced with respect to the environmental (or *I* band) potassium concentration. Sudden addition of excess potassium would involve readjustment in one direction, while addition of competing ions, particularly sodium, would also cause readjustment. Following the readjustment, a new stable structural equilibrium would appear, which might be nearly "normal" so far as contraction is concerned. In other words, the only effect of changes in intracellular potassium, so far as function of the tissue is concerned, might be to cause a transitory response during the shift of equilibrium. These considerations give a description of response of muscle to altered ionic environment that is identical with that detailed by Krogh *et al.* (1944). Treatment of the frog heart with small quantities of K-free Ringer's solution causes temporary cessation of beat, recovery taking place in a short time. Treatment with a fresh sample of K-free solution elicits the response again, a process that can be repeated many times. Each washing removes cellular potassium, yet each time the heart becomes readjusted to the altered content of intracellular ion. The sensitivity of cells to cations of the environment thus is not an indication of changes in the cation content of the cells, but of structural adjustment occasioned by upsetting an equilibrium.

From the recent work of Szent-Györgyi and others, and from the present discussion of cation localization, it is possible to build up a diagrammatic picture of the distribution of components in a muscle fiber. Two proteins, actin and myosin, must be considered. Actin, apparently, is the only protein known at present that could give directly the very long threads shown by the electron microscope (Schmitt) to extend through consecutive *A* and *I* bands. Myosin appears to exist in units

of about one micron long (Mommaerts, 1945). According to Buchthal and Knappeis (1943), there is somewhat less myosin than would be necessary to fill the *A* bands. Since the Szent-Györgyi group has shown a very strong bonding between myosin and actin, one could assume that a myosin-actin combination might tend to be more rod-like than either myosin or actin alone. If we further assume that there is some mutual attraction of myosin units, then the *A* bands could be represented as highly oriented actomyosin rods, the uncombined actin molecules extending on into the *I* bands, unsupported by myosin and, hence, less highly oriented and showing the characteristically lower birefringence. In the *I* bands, also, one would presumably find the globular proteins of muscle.

The tightly packed *A* bands may contain an excess of potassium, magnesium, and possibly some calcium. The *I* bands might be assumed to be in simple Donnan equilibrium with the external fluid environment, with respect to sodium, potassium, calcium, magnesium, and chloride. Calculations based on such assumptions as these lead to computed ion contents of whole muscle that fit published experimental results very closely.

If such a scheme is correct, then the major problem is to account for the distribution of ions between the *A* and the *I* bands, not between the external fluid and the muscle fiber. A part of this distribution probably relates to what can be called *ion binding*. According to several authors (cf., for example, Montigel, 1943), myosin selectively binds potassium. No explanation of this behavior has been offered. It may be that potassium, with its smaller effective diameter, can approach the charged groups of the protein closer than sodium. If the distance of approach were less than the critical 3.5 Å, association to a greater or lesser extent would take place. Again, the packing in the *A* bands might play a direct structural role in selection of K and exclusion of Na, the latter ion being highly hydrated and effectively larger.

The exchange resins may offer an instructive analogy here. These solid resins are well known as de-ionizers. According to figures cited by Myers (1943), at equal pH's, a sample of acid resin titrated with KOH will contain considerably more base than will a similar sample titrated with NaOH. In the cases of calcium and magnesium, one may invoke the well-known binding by protein in general as an explanation of concentrations observed. This may relate to the fact that, in spite of great hydration, these ions with their double charge present very concentrated centers of attraction.

This discussion has been based primarily on striated muscle. The same considerations of the relationships of cations and protoplasmic structure may, however, apply to other cells as well. The occurrence of particulate matter in all types of cells is now well known. The high degree of organization of striated muscle may be regarded as an extreme of such organization. Both granules and fibrillar structures of cells may well be related to cation distribution.

BIBLIOGRAPHY

Ashkenaz, E. W.
1938. Magnesium narcosis in muscle. J. Cell. & Comp. Physiol. **11**: 162.

Bailey, K.
1942. Myosin and adenosinetriphosphatase. Biochem. J. **36**: 121.

Banga, I.
1941–1942. The phosphatase activity of myosin. Studies **1**: 27.

Banga, I.
1943. Enzyme studies. Studies **3**: 64.

Barron, E., & S. Guzman
1943. An essay on comparative biochemistry. Adv. Enzymol. **3**: 149.

Boyer, P. D., H. A. Lardy, & P. H. Phillips
1943. Further studies on the role of potassium and other ions in the phosphorylation of the adenylic system. J. Biol. Chem. **149**: 529.

Boyle, P. J., & E. J. Conway
1941. Potassium accumulation in muscle and associated changes. J. Physiol. **100**: 1.

Bruman, F., & S. Finkelstein
1936. Die Wirkung kaliumreicher Futterung auf den Stoffwechsel von Ratten. Z. Exp. Med. **99**: 268.

Buchthal, Fritz, & G. G. Knappeis
1943. Correlation between changes in cross striation and mechanical tension in striated muscle fibre and their structural interpretation. Acta Physiol. Scand. **6**: 123.

Bureau, F.
1934. Recherches sur la répartition du potassium dans les cellules et sur les déplacements qu'il subit au cours des phénomènes d'excitation. Arch Int. Physiol. **39**: 311.

Caspersson, T., & Bo Thorell
1942. The localization of the adenylic acids in striated muscle fibers. Acta Physiol. Scand. **4**: 97.

Clark, A. J.
1938. The Metabolism of the Frog's Heart. Oliver and Boyd. London.

Conway, E. J., & G. Cruess-Callaghan
1937. Magnesium and chloride "permeations" in muscle. Biochem. J. **31**: 828

Deucher, F.
1941. Topochemische Untersuchungen uber Glykogen, Kalium und Aschegehalt in Warmblüterherzen. Z. Mikr. Anat. **49**: 401.

Dubuisson, M.
1942. Sur la répartition des ions dans le muscle strié. Arch. Int. Physiol. **111**: 439.

Engelhardt, W. A.
1941. Enzymatic and mechanical properties of muscle proteins (Translation). Adv. Cont. Biol. **14**.

Fenn. W. O.
1940. The role of potassium in physiological processes. Physiol. Rev. **20**: 377.

Fenn, W. O., & D. M. Cobb
1934. The potassium equilibrium in muscle. J. Gen. Physiol. **17**: 629.

Fenn, W. O., & D. M. Cobb
1936. Electrolyte changes in muscle during activity. Am. J. Physiol. **115**: 345.

Fenn, W. O., D. M. Cobb, J. F. Manery, & W. R. Bloor
1938. Electrolyte changes in cat muscle during stimulation. Am. J. Physiol. **121**: 595.

Fenn, W. O., & L. F. Haege
1942. The penetration of magnesium into frog muscle. J. Cell. & Comp. Physiol. **19**: 37.

Fenn, W. O., T. R. Noonan, L. J. Mullins, & L. Haege
1941. The exchange of radioactive potassium with body potassium. Am. J. Physiol. **135**: 149.

Gerendas, M.
1941–1942. Technisches über Myosinfaden nebst einigen Beobachtungen uber ihre Kontraktion. Studies **1**: 47.

Gersh, I.
1938. Improved histochemical methods for chloride, phosphate-carbonate and potassium applied to skeletal muscle. Anat. Rec. **70**: 311.

Greville, G. D., & H. Lehman
1943. Magnesium-calcium antagonism in muscle. Nature **152**: 81.

Hegnauer, A. H., W. O. Fenn, & D. M. Cobb
1934. The cause of the rise in oxygen consumption of frog muscles in excess of potassium. J. Cell. & Comp. Physiol. **4**: 505.

Heilbrunn, L. V.
1943. An Outline of General Physiology. W. B. Saunders. Philadelphia.

Heilbrunn, L. V., & P. G. Hamilton
1942. The presence of chloride in muscle fibers. Physiol Zool. **15**: 363.

Heppel, L. A.
1939. The electrolytes of muscle and liver in potassium-depleted rats. Am. J. Physiol. **127**: 385.

Heppel, A. L.
1940. Effect of age and diet on electrolyte changes in rat muscle during stimulation. Am. J. Physiol. **128**: 440.

Kalckar, H. M.
1942. The enzymatic action of myokinase. J. Biol. Chem. **143**: 299

Krogh, A., A.-L. Lindberg, & B. Schmidt-Nielson
1944. The exchange of ions between cells and extracellular fluid. II. Acta Physiol. Scand. **7**: 221.

Lissak, K., & T. Kovacs
1942. K and Ca-gehalt der Muskeln und Nerven in den verschiedenen Stadien der neuromuskularen Erregungsubertragung. Pflüg. Arch. **245**: 790.

Macollum, A. B.
1905. On the distribution of potassium in animal and vegetable cells. J. Physiol. **32**: 95.

Meigs, E. B., & W. G. Atwood
1916. The reactions of striated muscle to potassium chloride solutions. Am. J. Physiol. **40**: 30.

Miller, H. C., & D. C. Darrow
1940. Relation of muscle electrolyte to alterations in serum potassium and to the toxic effects of injected potassium chloride. Am. J. Physiol. **130**: 747.

Miller, H. C., & D. C. Darrow
1941. Relation of serum and muscle electrolyte, particularly potassium to voluntary exercise. Am. J. Physiol. **132**: 801.

Mond, R., & H. Netter
1932. Über die Regulation des Natriums durch den Muskel. Pflüg. Arch. **230**: 42.

Montigel, C.
1943. Myosin und Kalium. Helv. Physiol. Acta **47**: 1.

Mommaerts, W. F.
1945. Molecular morphology of myosin. Nature **156**: 631.

Moog, F., & H. B. Steinbach
1945. Adenylpyrophosphatase in chick embryos. J. Cell. & Comp. Physiol. **25**: 133.

Myers, R. C.
1943. Adv. Coll. Sci. **1**: 317.

Sandow, A.
1945. The mechanism of energizing muscular contraction. Trans. N. Y. Acad. Sci. (Ser. II) **7**: 139.

Scott, G. H., & D. M. Packer
1939. An electron microscope study of magnesium and calcium in striated muscle. Anat. Rec. **74**: 31.

Spiegelman, S., & H. B. Steinbach
1945. Substrate-enzyme orientation during embryonic development. Biol. Bull. **88**: 254.

Steinbach, H. B.
1940a. The distribution of electrolytes in *Phascolosoma* muscle. Biol. Bull. **58**: 444.

Steinbach, H. B.
1940b. Electrolyte balance of animal cells. Cold Spring Harbor Symposia **8**: 242.

Steinbach, H. B.
1940c. Potassium in frog muscle. J. Biol. Chem. **133**: 695.

Steinbach, H. B.
1944. The osmotic behavior of frog sartorius muscles. J. Cell. & Comp. Physiol. **24**: 291.

Szent-Györgyi, A.
1941–1942. The contraction of myosin threads. Studies **1**: 17.

Szent-Györgyi, A.
1943. Chemistry of muscle. Bull. Soc. Chem. Biol. **25**: 242.

Tabor, Herbert, & S. M. Rosenthal
1945. Electrolyte changes in tournequet shock. Publ. Health Rep. **60**: 401.

Verzár, F.
1942. Muskelermudung und Nebenniere. Schweiz. med. Wochenschr. **72**: 661.

Wilde, W.
1945. The chloride equilibrium in muscle. Am. J. Physiol. **143**: 666.

Wood, W. H., & G. K. Moe
1942. Electrolyte and water content of the ventricular musculature of the heart-lung preparation with special reference to the effects of cardiac glycosides. Am. J. Physiol. **136**: 515.

Yannet, H. & D. C. Darrow
1940. The effect of depletion of extracellular electrolytes on the chemical composition of skeletal muscle, liver, and cardiac muscle. J. Biol. Chem. **134**: 721.

Zwikster, G. H., & T. E. Boyd
1935. Reversible loss of the all or none response in cold blooded hearts treated with excess potassium. Am. J. Physiol. **113**: 560.

THE INTERACTION OF MYOSIN AND ADENOSINETRIPHOSPHATE

By Shih-Chang Shen

The Institute of Physiology and Psychology, Shanghai, China

The unique structure of muscle has long been a subject of intensive study and speculation. It is the first and perhaps the only biological structure in which the fundamental relationship between morphological form and protoplasmic organization has been clearly demonstrated. The elongated muscle tissue is essentially a bundle of long protein fibers arranged parallel to each other. The latter are, in themselves, made up of elongated protein particles, myosin, grouped together in a more or less end-to-end fashion. That the myosin particles are the basic structural units of muscles, has been a well accepted view. Although it seemed only logical that these structural units might also be the functional elements of the muscle, the surmise has not, until recently, been substantiated by any direct evidence. In other words, while a striking similarity in basic structure of a muscle fibril to that of a myosin fiber was observed, there was no evidence to show that the latter was capable of activities analogous to that of the muscle. The complete lack of such evidence, heretofore, must be regarded as the chief handicap in the postulation of a plausible mechano-chemical theory of muscular activity.

In 1939, Engelhardt and Ljubimova announced their startling finding that myosin is itself an enzyme, adenosinetriphosphatase (ATPase). The important point of departure was thus established, that myosin can no longer be regarded as an inert protein serving merely as the building blocks of the muscle fibril, but is, in all probability, a functional basis for the muscular activity. The importance of this finding lies further in the following facts.

(1) Among the processes of intermediary metabolism in the muscle, the nearest in time to the contraction of the fibrils is the breakdown of adenosinetriphosphate (ATP) to adenosinediphosphate (ADP) and inorganic phosphate.

(2) In the muscle metabolism, this is the exothermic reaction that is by far the most likely to be capable of supplying the bulk of the free energy for muscular activity. Although some ATP is broken down by the transfer of phosphate to hexosemonophosphate (Neuberg ester), a transfer which probably liberates free energy, the quantitatively largest

part of its breakdown occurs by the splitting-off of free phosphate under the influence of the enzyme ATPase.

It became at once apparent that a causal relationship between the chemistry and mechanics of muscles may be sought in the interaction of myosin, as an ATPase, and its natural substrate, ATP. This paper attempts to make a brief review of the experimental observations by various investigators along this line, and to discuss their general bearing on muscular activity.

GENERAL PHYSICAL CHARACTERISTICS OF MYOSIN

The elongated, or anisometric, character of the particles of myosin was first suggested by Brücke (1858), and demonstrated by the classical experiment of von Muralt and Edsall (1930). By virtue of the elongated, rod-like or even thread-like, structure of its molecules, myosin has the extraordinary property of forming fibers of considerable strength. Weber (1934) first produced myosin fibers which were shown to possess many of the properties of muscle fibers. Myosin fibers can be prepared very simply by squirting a solution of myosin through a fine capillary into a large volume of distilled water wherein the myosin sol coagulates in the form of a fine thread These threads have appreciable tensile strength, being capable of supporting their own weight up to a length of a meter, even though they contain only about 1 per cent protein. Upon drying under slight tension, the fibers show strong positive double refraction and high tensile strength. Further stretching of the fibers, after slight moistening, appears to bring about nearly complete orientation of the protein particles along the length of the fibers, which then exhibit an x-ray diffraction diagram virtually identical with those of muscle. Equally interesting is the optical behavior of these fibers. Immersed in fluids of varying refractive index n_2, after complete penetration of the immersion fluid into the interior of the fibers, they give an n_2-double refraction curve very similar to that of the intact muscle fiber. In both cases, the double refraction at the minimum is quite large, indicating that, in addition to the structural birefringence due to parallel arrangement of the myosin molecules in the fiber, there is an intrinsic birefringence, due to the anisotropic nature of the myosin molecules themselves

In view of these striking similarities between the myosin fiber and the muscle fibril, the contention may well be justified that a properly prepared myosin thread is, in effect, a molecular model of the muscle fiber. It is, therefore, of great interest to see if such a model would

not also approximate the mechanical behavior of an intact muscle. Having discovered the enzymic property of myosin, and that the ATPase activity was well preserved even in myosin threads, Engelhardt and his associates set out to investigate the interaction of ATP and myosin in relation to the mechanical properties of myosin threads

EFFECT OF ATP ON MYOSIN THREADS

For the investigation of the mechanical properties of myosin threads, Engelhardt used the apparatus as shown in FIGURE 1.

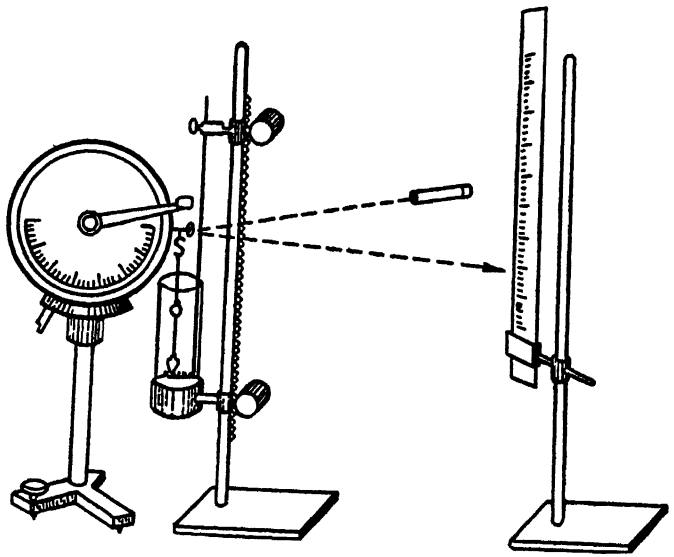


FIGURE 1 Arrangement of apparatus for testing the mechanical properties of myosin threads

The setup consists of a torsion balance, to the lever of which one end of the myosin thread is attached. The other end of the thread is so fixed that the thread is stretched vertically under a minimum tension. The entire length of the thread is kept immersed in a vessel containing KCl (serving as a control), into which various chemical substances to be tested can be added. When a load is applied to the thread, a change in the balance reading results, giving the magnitude of the load The change in length of the thread is measured by a light beam reflected on a scale, calibrated in millimeters, by means of a mirror attached to the balance lever. The extensibility of the thread under a constant load serves as an index of its mechanical properties.

A number of compounds which might play a part in muscle metabolism are investigated for their effect on the myosin thread. Of the substances thus far tested (ATP, ADP, adenylic acid, and inorganic phosphate), only one shows any noticeable effect upon the myosin thread This only exception is ATP, which is found to increase the extensibility of the myosin thread by from one and one-half to two times. The result is graphically presented in FIGURE 2.

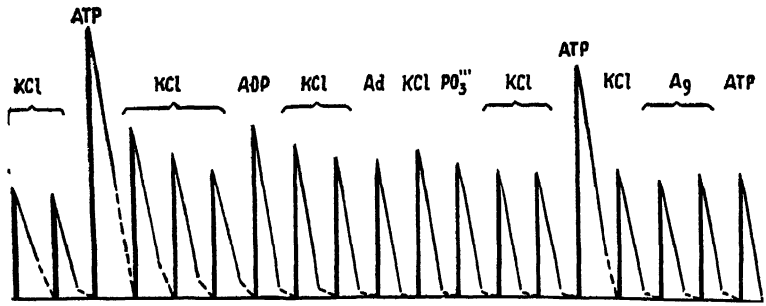


FIGURE 2. Load constant, — 200 mg ATP = adenosine-triphosphoric acid; ADP = adenosine-diphosphoric acid; Ad = adenylic acid; PO ''' = phosphate solution, Ag = $AgNO_3$

Since myosin has been shown to be a specific enzyme for the breakdown of ATP, the question immediately arises: Is the mechanical change in myosin thread directly attributable to the enzyme-substrate reaction? If so, it follows that when the ATPase activity of myosin is blocked by some means, the addition of ATP would produce no effect on the myosin thread. That such is indeed the case is suggested by the following facts:

1. Changes in extensibility as induced by ATP are observed only with relatively fresh myosin threads. Threads which have been kept for several days may retain their normal extensibility without any marked change, but their response to the action of ATP gradually disappears. As a rule, this occurs simultaneously with the loss of the enzymic properties of the fiber.

2. It is found that the ATPase activity of myosin is completely and specifically destroyed by silver salts in extremely small concentrations—of the order of 1:100,000 M. In the presence of, or after preliminary treatment with, silver, the myosin threads lose their ability to change their mechanical properties under the action of ATP. The silver in no way affects the normal extensibility of the myosin threads.

3. Thermal treatment (exposure to 40° C.) of the threads, which destroys their enzymic properties, also eliminates the effect of ATP

on their extensibility. However, in this case, the normal extensibility of the threads is slightly changed. In this connection, it is necessary to emphasize the fact that, contrary to the general expectations, no characteristic changes occur in the hydrogen ion concentration within the physiological zone, pH 5.0 to 7.5. Outside this zone, the threads disintegrate, tear, and break.

Another question of fundamental importance is: In what manner are the mechanical properties of myosin threads affected by the breakdown of ATP? For the present, we need not be concerned with the actual continuity of the two phases, contraction and relaxation, in a normal muscle curve. The point under question is whether the immediate effect of ATP breakdown is a lengthening or a shortening of the myosin particles. The result obtained by Engelhardt and coworkers seems to indicate clearly that the former is true, namely, the breakdown of ATP affects the myosin thread, not in the sense of contraction but of relaxation. A fuller discussion on this point will be taken up later, when similar investigations by other workers have been considered.

BEHAVIOR OF MYOSIN SOLUTIONS

The use of myosin threads as a mechanical model for muscle fibril has a distinct advantage in that the changes in the mechanical properties of such threads can be easily observed and directly measured. It does not, however, give clear-cut evidence of the behavior of individual myosin molecules or micelles. In a sufficiently diluted solution of myosin, the protein particles may be assumed to remain in a freely orientable state. At any rate, complications arising out of the intermicellar reactions are reduced to a minimum, thereby facilitating a more proper evaluation of the mechanical behavior of the individual particles. One obvious difficulty here is the lack of means for a simple and direct measurement of the changes in these particles. There are, however, a number of indirect methods for the investigation of molecular anisometry. Among these, perhaps the most promising approach at the present time is the method of simultaneous measurement of anomalous viscosity and flow-birefringence in a modified form of the coaxial (Couette) viscosimeter. The theory and practice of this method have been discussed in detail by von Muralt and Edsall (1930), Robinson (1939), Lawrence, Needham and Shen (1944), and in the excellent review of Edsall (1942). It is only necessary here to mention briefly the general principles of the method.

A solution or suspension of elongated particles, when subjected to

varying shear forces, exhibits the following properties: (1) There is an anomalous dependence of its viscosity on shear rate. (The viscosity of a Newtonian fluid is independent of the shear force.) Its relative viscosity is markedly higher than normal solutions. In polymers, the relative viscosity is found to be proportional to the polymer length or axial ratio at equal molar concentrations. (2) The solution shows a well-marked flow-birefringence, although it is completely isotropic when at rest. Both the intensity of flow-birefringence and the "angle of isocline" (degree of orientation of the particles along the streamline of flow) are measures of particle asymmetry, other factors remaining constant. It must be admitted, however, that there is as yet no entirely satisfactory prescription for an accurate estimation of the absolute dimensions of the particles by this method alone. Nevertheless, the method is adequate, at least, to determine the relative change in particle shape as affected by various agents.

Using this method, Needham and co-workers made a detailed study of the myosin solutions and the effect of ATP and related substances (Dainty *et al.*, 1944). For the present discussion, it suffices to give a summary account of their observations. When a myosin solution is treated with ATP, its flow-birefringence is markedly decreased, its anomalous viscosity is retained, and its relative viscosity is decreased. This is interpreted as a shortening or reduction in the axial ratio of the myosin particles, as a result of ATP action. Although a variety of electrolytes, such as Li^+, Na^+, K^+, NH_4^+, Ca^{++} and Mg^{++}, also reduce to some extent the flow-birefringence of myosin solutions, their effect on its viscosity varies. Some cause an increase and others a decrease. Moreover, the amount of these electrolytes required to produce a marked reduction in myosin flow-birefringence is far greater than that of ATP. For a myosin solution of less than 1 per cent protein, approximately the full effect of ATP is obtained at 0.004 M, falling off very rapidly at lower concentrations, and being very little higher in amounts up to 0.2 M. To produce an effect comparable to that caused by 0.004 M ATP, it would require 1.5 M urea, 1.0 M monovalent cations, and 0.2 M divalent cations: or a pH of 10. The high potency of ATP in reducing the particle asymmetry of myosin is taken, together with other evidence, to indicate the specificity of the ATP effect. Variations of pH between 6.0 and 8.5 have no appreciable effect on the flow-birefringence or the relative viscosity of myosin, and this finding is in general agreement with Engelhardt's observation on myosin threads. Perhaps the most significant result of this investigation is the observation that changes in the physical properties of

myosin, brought about by ATP, are spontaneously reversible. This is in contrast to the findings of Edsall and Mehl (1940) on the effect of electrolytes which cause an irreversible reduction in the flow-birefringense of myosin. Following their initial and apparently instantaneous drop, upon the addition of ATP, both the flow-birefringence and relative viscosity of a myosin solution return simultaneously to their original values, over a length of time depending on the temperature and the purity of the myosin preparation. Thus, the recovery time is much shorter at 37° C. than at 20° C.; and a thrice precipitated myosin recovers much sooner than does a once precipitated sample. The temperature effect is related to the rate of ATP hydrolysis. The purity of myosin preparation is suggestive of the absence of myokinase which tends to replenish the substrate by converting two molecules of ADP to one of ATP and one of adenylic acid (Kalckar, 1942). The reversibility of the physical changes in myosin as a result of ATP treatment is therefore shown to be closely related to the enzymic splitting of ATP by myosin. One of the most demonstrative of such experiments is graphically represented in FIGURE 3, which shows three successive falls and returns of flow-birefringence in the same sample of myosin upon three successive additions of ATP; two successive falls and returns of relative viscosity; together with the course of the ATP breakdown over the first period. It should be noted that, although the completion of the recovery both in flow-birefringence and in relative viscosity is attained after a relatively long period, recovery follows, and proceeds rather rapidly at first, immediately after the initial reduction, which is completed in a very brief time. In fact, the recovery process sets in so rapidly, especially in the absence of myokinase, that it is almost impossible to determine the lowest point reached by the drop of flow-birefringence after the addition of ATP. Another observation, which is not shown in the graph but should be mentioned, is that, along with the fall and return of flow-birefringence and relative viscosity of the myosin solution, there is a corresponding decrease and recovery in the angle of isocline. This evidence strongly indicates a genuine change in the axial ratio of the myosin particles, rather than a photoelastic effect. Nor can the change in flow-birefringence in myosin solution be due to a change in its refractive index as a result of the addition of ATP or other reagents. Direct measurement reveals no appreciable change in the refractive index.

In addition to ATP and a variety of neutral salts, a number of intermediate compounds in carbohydrate breakdown and substances related to them have also been examined for their effect on myosin solu-

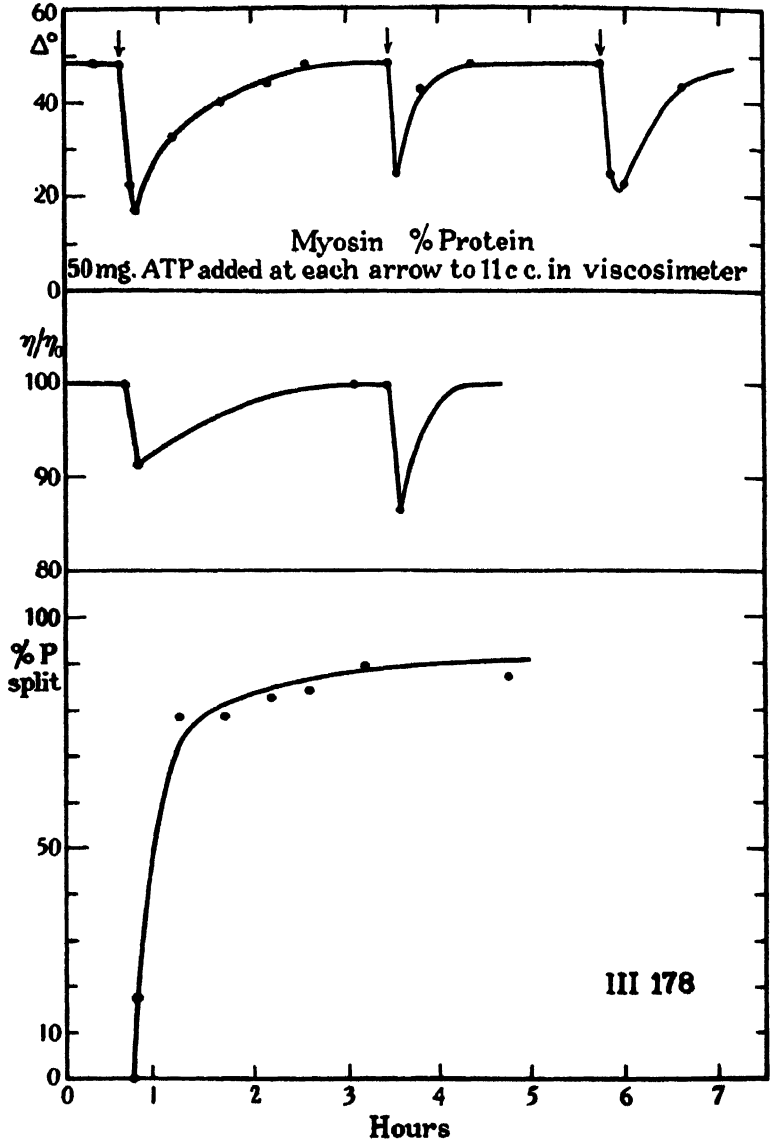


FIGURE 3 Three successive falls and recoveries of flow-birefringence of a myosin sol treated three times with adenosinetriphosphate, with two successive falls and recoveries of relative viscosity (expressed in percentage of initial value), and estimations of inorganic phosphate liberated by adenosinetriphosphatase activity of the myosin during the first cycle

tions. These include, besides various forms of inorganic phosphates, adenosinediphosphate, adenylic acid, inosinetriphosphate, and inosinic

acid. Effects similar to those of ATP on the physical properties of purified myosin have been obtained, so far, only with inosinetriphosphate, which is also split by myosin into IDP and inorganic phosphate. It is of interest to note that, while inorganic triphosphate is also split by 1- to 2 per cent solution of three-times precipitated myosin, it does not affect the flow-birefringence or viscosity of the myosin. It seems, therefore, plausible to say that the purine-ribose end of the nucleoside-triphosphate molecule is of much importance in changing the shape of the myosin molecule or micelle. This view is shared by Szent-Györgyi's school in their recent work on actomyosin, which will be presently discussed.

BEHAVIOR OF ACTOMYOSIN

A great deal of new and important information on the muscle proteins has been provided by the recent work of Szent-Györgyi and his associates (Straub, 1942; Banga and Szent-Györgyi, 1942). Their work has been reviewed and a theory of muscular contraction advanced by Szent-Györgyi (1946). Only a brief account of their observation is given below.

The most significant feature in the investigation of the Hungarian school is the discovery that what has hitherto been referred to as "myosin" is, in effect, a complex of two separable water-soluble proteins: actin and myosin. The two proteins reversibly combine to form actomyosin. The old "myosin" is therefore to be regarded as actomyosin of indefinite composition of the two compound proteins. When not in combination, neither of the two proteins, actin and myosin, is markedly elongated in its particle shape. Nor will either alone form threads. Actin, however, is capable of existing either in its globular form (G-actin) or in fibrous form (F-actin). The physical properties of actin and myosin are highly susceptible to changes in the cationic environment, such as Na^+, K^+, Ca^{++}, and Mg^{++}. Thus, myosin readily binds cations, having a higher affinity for bivalent than monovalent ions. It is precipitated by KCl at concentrations higher than 0.001 M, maximally at 0.025 M; but redissolved at 0.1 M. The transformation of G-actin to F-actin is dependent on the presence of K^+ and traces of Mg^{++}. The process is completely inhibited by traces of Ca^{++}. Whereas ATP has no effect on actin, it is readily bound by myosin. If the quantity of ATP bound by the unit weight of myosin is plotted against the ATP concentration, a straight line is obtained, suggesting that it is some sort of a "chemical" link which joins the two substances.

The enzymic activity of ATPase is found only in the myosin moiety of actomyosin, but not in actin.

The general physical properties of the complex actomyosin are much the same as those described for "myosin" by earlier workers. Thus, actomyosin solutions show high and anomalous viscosity, strong flow-birefringence, and can readily be spun into threads. Like all highly polymerized fibrous colloids, actomyosin forms a thixotropic gel at high concentrations. It is highly hydrophilic. In the absence of salts, it swells to a very great extent. The swelling is prevented by 0.001 M KCl, or by other neutral salts. If the salt is added to a swollen gel, this becomes turbid and shrinks. Like myosin, actomyosin is also precipitated by small concentrations of neutral salts. At 0.05 M KCl, maximum precipitation of actomyosin occurs. Upon raising the KCl concentration to above 0.05 M, the precipitate decreases in solidity and liquefies at 0.4 M. On further increase of the salt concentration, the viscosity gradually drops and reaches a minimum at 2.0 M, corresponding to the additive value of the viscosity of myosin and actin, indicating that the actomyosin has completely dissociated into its components. Perhaps the most instructive effect observed of the observation is the remarkable influence of ATP on the colloidal behavior of actomyosin. The presence of small amounts of ATP profoundly changes the solubility-KCl curve, as illustrated in FIGURE 4.

The range of the salt concentrations within which actomyosin first

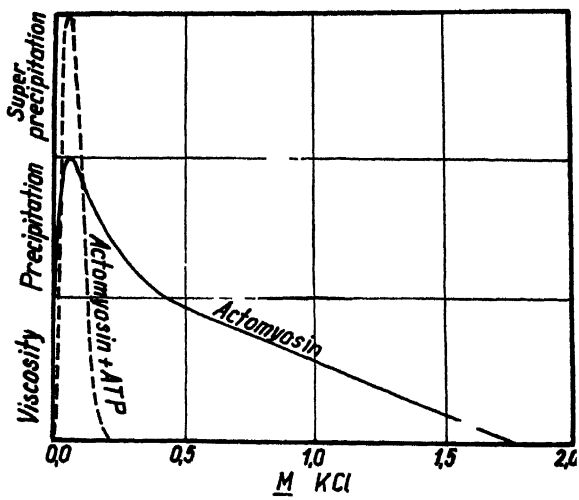


FIGURE 4. Semi-quantitative curve showing the behavior of actomyosin in presence and absence of ATP at various KCl concentrations.

gels, then redissolves, and finally completely dissociates, is greatly reduced by ATP. Thus, actomyosin is completely dissociated at 0.2 M, in the presence of ATP; as compared with 2.0 M, in the absence of ATP. It is further shown that, in the absence of ATP, the precipitate formed by actomyosin appears rather loose and hydrophilic; while, in the presence of ATP, the precipitate is very coarse, granular and contains little water. The latter phenomenon is termed by the investigators as "super-precipitation" to distinguish it from the simple salt precipitation. Identical in nature with the super-precipitation of actomyosin suspensions is the rapid and excessive shrinkage of a salt-free actomyosin gel, upon the addition of salt, in the presence of ATP. The effect of ATP is most dramatically demonstrated on actomyosin threads. An actomyosin thread is prepared in a 0.05 M KCl solution where the protein is insoluble. If the thread is now treated with ATP, it contracts within a few seconds to about one-third of its original length, and becomes proportionately thinner. The contraction is, however, not spontaneously reversible. But recovery can be induced by altering the ionic environment.

When dissociated actomyosin is caused to contract, the reaction takes place in two distinct steps: the union of actin with myosin to form "relaxed actomyosin," and the contraction of the "relaxed actomyosin" to form "contracted actomyosin." The reaction can be perfectly reversed by shifting the temperature, the equilibrium being shifted to the relaxed form at low temperature, and to the contracted form at room or body temperature. The passing of relaxed into contracted actomyosin can also be brought about by slight variations (15 per cent) of the concentration of either ATP or the salt, with one of the two remaining constant. The ATP saturation of the protein depends on the ATP concentration present. To have a strongly reactive system, it requires a high ATP saturation and, hence, high ATP concentration. If the ATP concentration is decreased, it is the relaxation that suffers first and, consequently, the system begins to contract. The reaction of actomyosin depends also on the relative concentration of the actin and myosin that make up the complex. The two proteins can react in all proportions, but the most reactive system is one which has an actin-myosin ratio of 2.5. This is found to be the composition of actomyosin in the normal muscle, which thus contains maximally active actomyosin.

It is necessary to emphasize here that all the above descriptions of the physical behavior of actomyosin apply only to the complex formed by myosin with F-actin, but not with G-actin. Although G-actin also

readily forms a complex with myosin, the resulting actomyosin (G-actomyosin) has quite different physical properties from that of the F-actomyosin. G-actomyosin has a low viscosity, about the same as that of free myosin. It does not form gels or threads. It is, therefore, not contractile, with or without the help of ATP. The remarkable fact is that, in the presence of ATP, G-actomyosin dissociates at any salt concentration into G-actin and myosin or, more correctly, into G-actin and myosin-ATP. Hence, in order to form an active actomyosin that is reactive to ATP, the G-actin must be first transformed into F-actin, either before or after its combination with myosin. This reaction, as mentioned before, is catalyzed by traces of Mg^{++}, and completely inhibited by Ca^{++} in equally small amounts.

THEORIES OF MUSCULAR CONTRACTION-RELAXATION

All mechano-chemical theories of muscular activity based on the physical behavior of myosin start with the assumption that the changes in the mechanics of muscle are, basically, the result of the corresponding changes in the myosin molecules or micelles. So far, there seems to be little objection to this assumption. The task, then, remains one of establishing a link between the chemical and mechanical changes of the contractile elements. Is this link a direct one or an indirect one? The question was very clearly put by Bernal (1938): "Do the chemical reactions (in muscle) have to do directly or indirectly with the actual contraction process? If the first is true, they are probably of a topochemical nature, that is, reactions in which one of the reagents is not a free molecule but a radical bound in a fixed position on a chain molecule. On the other hand . . . the chemical reactions may bring about contraction indirectly by building up a potential field, probably of an electrical nature, in which the chemically inert myosin molecule contracts reversibly." Recent experimental evidence all seems to support the view that the first alternative is the more likely one. Especially significant is the discovery by the Engelhardt group that myosin is *not* a chemically inert protein, but endowed with a highly potent enzymic activity which is responsible for the most energy-rich reaction in the metabolic processes of muscle. On the other hand, the evidence for the identity of myosin and ATPase, hitherto resting upon the failure of all attempts to separate them, must remain of a negative character. At any rate, there is still no evidence to show that, under physiological conditions, myosin-ATPase may reversibly dissociate to form free myosin and free ATPase.

Indeed, experimental evidence all seems to indicate an unusually strong bond between myosin and ATPase, *in vivo* or *in vitro*. The enzyme is so firmly attached to myosin that it can virtually be regarded, under all normal conditions, as an integral part of the contractile protein. In any case, whether or not ATPase is separable from, though firmly bound to, myosin, it does not materially affect the immediate bearing of the deep-seated effect of ATP on the physical state of myosin or actomyosin, in relation to muscular contraction.

A casual comparison of the results of the three groups of investigations outlined in the preceding pages may reveal a serious discrepancy among them. Thus, Engelhardt's group, while failing to detect any contraction of the myosin threads, clearly demonstrated that ATP causes the myosin threads to extend, corresponding to the recovery relaxation in muscle. On the other hand, investigations both on myosin solutions and actomyosin threads seem to show that the immediate effects of ATP are, respectively, a shortening of the myosin micelles and a contraction of the actomyosin threads. Upon closer examination, it is soon seen that the contradiction in results is merely an apparent one. First of all, it is necessary to take into consideration the difference both in experimental techniques and in materials used in the three different groups of investigations. The difference in techniques can account for a possible difference in degree, but not in direction. More significant is the difference in the experimental material. Although both Engelhardt's and Szent-Györgyi's groups used myosin threads, the threads were prepared and subsequently treated in rather different manners with regard to the salt concentrations. In view of the fact that slight variations in the salt concentration may produce a rather marked effect on the physical properties of myosin threads, it is conceivable that the threads might have behaved somewhat differently in the two schools of investigations. Moreover, there is the possible difference in the actin-myosin composition in the "myosin" and actomyosin threads. The conditions are, again, very different from those for threads in the contraction of the myosin particles in a dilute solution subjected to a constant shear. Nevertheless, in theory or in practice, none of these differences in experimental conditions suffices to explain the diametrical contrast, if such be the case, in the results. As a matter of fact, despite the considerable difference in technique and in material, these investigations are in fundamental agreement in their observations. They all seem to agree not only on the general observation that ATP has a profound and specific effect on the physical behavior of myosin, but specifically on the fact that it is the re-

covery-relaxation of myosin that is closely associated with, and probably directly attributable to, the enzymic breakdown of ATP by myosin. This relation is clearly shown in FIGURE 3. Practically the entire course of ATP hydrolysis takes place during, and proceeds in close parallelism to, the recovery relaxation of myosin. No less clear-cut an evidence, although of a somewhat indirect nature, is the observation of Engelhardt's school that, when the ATP breakdown is completely inhibited by traces of silver, no increase in extensibility is produced by ATP on myosin threads. From the standpoint of muscle energetics, such a mechano-chemical relationship seems at least theoretically sound. Since a large supply of energy is essential for the relaxation and recharging of the myosin fibril, the most probable source for the bulk of free energy would be from the splitting of ATP.

The question now arises as to how we account, then, for the actual contraction of myosin particles or threads as a result of ATP treatment? In line with the general two-affinity theory of enzyme action, the myosin-ATP reaction involves two distinct steps in the order of: (1) the formation of a myosin-ATP complex, and (2) the actual breakdown of ATP by the enzyme. As shown above, it is the second step in the enzyme-substrate reaction that is closely associated with the recovery-relaxation process of myosin. Whether a small part of the ATP breakdown is also related to the contraction phase of myosin is not clear from the available data. If we regard the ATP breakdown primarily as a source of free energy, we may then ask whether the actual contraction of muscle or myosin requires external energy. If, on the one hand, a resting muscle is to be compared with a charged condenser or a piece of stretched rubber, then it would contract, when initiated by a "trigger" action, merely by passing from a thermodynamically less stable level to a more stable low energy level. The difference in energy resulting from contraction performs work in isotonic contraction, and is mainly dissipated as heat in isometric contraction. If, on the other hand, the process of contraction involves an active compression, or "supercontraction," of a passively extended molecule, external energy would necessarily be required. A choice of the two alternatives must, however, await further investigation. In the experiment on myosin threads by the Engelhardt group, no contraction or a decrease in extensibility of the thread was observed, under all conditions. Without experimental confirmation of their observation, it can only be surmised that their failure to detect any contraction may be due to the inadequate sensitivity of the technique used, or, as is much more likely, that the contraction phase (which is an ex-

tremely rapid process) of the threads in the presence of ATP is completed and relaxation sets in before the measurement of thread extensibility can be taken. As shown in the experiments with myosin solutions, especially where the protein has been purified by repeated precipitation, the contraction of myosin particles caused by ATP is such a rapid and transient response that it is often difficult to determine the maximum shortening of the particles. Only with techniques much improved beyond those hitherto used may we hope to establish an exact time-relationship between the contraction phase and the onset of ATP breakdown. The investigators then resorted to an indirect approach, by attempting to block the ATPase activity and to ascertain thereby the dependency of the contraction process on the enzymic reaction. The destruction of the ATPase activity of myosin was first tried by temperature treatment (20′ at 37° C.). It was, however, found impossible to inactivate the enzyme by such a treatment without, at the same time, greatly altering the physical state of the protein solutions. A similar experience with myosin threads was also encountered by the Engelhardt school. The result, therefore, was not conclusive. Based on the finding (Needham, 1942; Bailey, 1942) that Ca^{++} greatly activates ATPase, the Cambridge investigators further tried the effect of ATP on oxalated myosin. This attempt again met with failure, as it was found that oxalate does not prevent splitting of ATP, when added to Ca-myosin. It should be emphasized, at this point, that, even if a potent enzyme inhibitor could be found which would completely block the ATPase activity and, at the same time, prevent the contraction of myosin, such a discovery would not rule out the possibility that the inhibition is due to the prevention of a normal enzyme-substrate complex formation. In fact, many enzyme inhibitors are known to exert their effect by competing with the normal substrate in their affinities for the enzyme. The following experiment by Needham's group is, therefore, an instructive one. They found that inorganic triphosphate, though ineffective in changing the shape of myosin particles, is split by purified myosin. It must, therefore, have a certain affinity for myosin, without affecting the physical properties of the protein. In the presence of sodium triphosphate, the effect of ATP in reducing the flow-birefringence and relative viscosity of myosin is markedly inhibited, although the amount of inorganic P formed by the enzyme action of myosin is considerably higher than when either ATP or sodium triphosphate alone is present. This evidence is strongly in favor of the view that the actual contraction of myosin is closely associated with the formation of a specific enzyme-substrate

complex, myosin-ATP, and is independent of the subsequent breakdown of ATP. This latter process, as we may recall, is related to the recovery-relaxation of myosin.

Thus far, we have, at best, merely established a time-relationship between the chemical and mechanical events in the interaction of myosin and ATP. The exact manner in which myosin molecules or micelles change in length remains entirely obscure. At present, we are in complete ignorance as to whether such changes are micellar in character or whether they are analogous to the configurational changes in the molecules of keratins. Much light has been shed on this problem by the brilliant work of Szent-Györgyi and his associates on actomyosin and its components. According to this school, the contraction and relaxation of actomyosin can be explained adequately by its colloidal behavior. The interpretation is therefore essentially on a micellar, rather than molecular, level. The contraction of actomyosin is, thus, a process of dehydration and "superprecipitation" of the protein, under appropriate conditions. The relaxation is one of hydration, solvation, and eventually complete dissociation of the contracted actomyosin (FIGURE 4). On the basis, largely, of these observations, we may formulate the following relatively simple scheme of the mechanochemistry of muscular activity.

The essential feature in the contractile elements of a resting muscle is the complete dissociation of actomyosin. This condition is maintained so long as actin remains only in its globular form. G-actin is itself non-contractile, and is incapable of forming active actomyosin with myosin or myosin-ATP. The resting muscle, to be a potentially reactive system, must also contain a high level of ATP, such as is normally found in the muscle. It is not certain, however, whether, at this stage, myosin exists as free myosin or as myosin-ATP complex. Except when a "partition" could be shown to exist between myosin and ATP, the two readily combine to form a complex. In either case, the muscle will remain relaxed, since neither myosin nor myosin-ATP would associate with G-actin to form a contractile actomyosin. However, if myosin is present as myosin-ATP, an explanation must then be sought for the inactivity of the ATPase in a resting muscle. According to Banga, the ATPase activity of myosin is greatly inhibited, while that of actomyosin is activated, by traces of Mg^{++}. By accepting this explanation, we meet with another difficulty. That is, the presence of Mg^{++} in a resting muscle should activate the transformation of G-actin into F-actin, and the latter, by its subsequent reactions, would throw the muscle into contraction. This difficulty may be got around

by introducing another variable, namely, K^{++}. The association of G-actin to form F-actin occurs, in the presence of Mg^{++}, only at a certain K^+ concentration. Summing up the alternatives, we may then assume that, for a resting muscle to pass into contraction, the process involves a shift in the concentration of either K^+ or ATP. This is in line with the *in vitro* experiment, where it is shown that dissociated actomyosin can be brought to association at a constant ATP, by varying the salt concentration; or, at a constant salt concentration, by varying the ATP concentration. Very slight variations (15 per cent) suffice to produce these changes. It seems that, according to these assumptions, there should be, in the resting muscle, some sort of partitions between the electrolytes and ATP, with respect to the contractile proteins, their interactions being regulated by the breakdown of such partitions during contractions. The possibility of such a spatial arrangement in the muscle of the electrolytes and ATP may find its support in histochemical indications that the metal ions such as K^+, Ca^{++} and Mg^{++} seem to be preferentially located in the anisotropic regions of the resting muscle fibril (Dubuisson, 1942); while the adenylic derivatives, presumably ATP, are found to be localized in the isotropic regions (Caspersson and Thorell, 1942). The manner in which these substances are regulated during activity is not known, apart from the fact that during contraction there is a "mixing" of the two regions along the muscle fibril.

Upon stimulation, a shift in the concentration of either K^+ or ATP results. In the presence of Mg^{++} and at a proper concentration of K^+, G-actin is converted into F-actin. F-actin then combines with the myosin-ATP complex, and passes directly into contracted actomyosin. Alternatively, F-actin may first combine with free myosin to form relaxed actomyosin which, in the presence of ATP, becomes contracted, resulting in a contraction of the muscle. In its contracted state, actomyosin begins to split ATP, being ativated by Mg^{++} which has hitherto held in abeyance the ATPase activity of myosin. As a result of the depletion of ATP, actomyosin dissociates into relaxed actomyosin and, eventually, into actin and myosin. The dissociation of actomyosin, as well as that of F-actin, is probably also catalyzed by Ca^{++} which may be brought into action in the process of contraction. It is interesting to note that, in the actomyosin model constructed by Szent-Györgyi, F-actin is forcibly broken up into G-actin by the shortening of the myosin part of the complex. The breakdown of ATP, meanwhile, supplies the necessary energy for the relaxation of the muscle, as well as for the resynthesis of ATP, thus completing the cycle. The ATPase

activity automatically stops when actomyosin becomes relaxed and dissociated.

The above, then, is an outline of a mechano-chemical scheme in its simplest form of muscular contraction and relaxation. The picture is, of course, far from complete, leaving many points open to question and further investigation. On the "mechanics" side, the changes in the micellar character of the muscle proteins during activity may be related to possible changes in the protein molecules. Both Needham's and Szent-Györgyi's observations give evidence to emphasize the role of the purine-ribose end of the nucleoside-triphosphate in bringing about the changes in the shape of the protein molecules. As stated by Szent-Györgyi: "To be more exact, it is not the double protein (actin-myosin) which contracts and relaxes, but its nucleotide compound." The effect of electrolytes seems to be, essentially, on the colloidal state of the protein, presumably by altering the intermicellar forces, resulting in hydration (relaxation) or gelation (contraction) of the protein. The effect of ATP, on the other hand, seems to be more deep-seated. It probably alters the intra-molecular groupings to produce a folding or unfolding of the protein chains. It would be idle to discuss these possibilities further at present. Much may be expected from the x-ray analysis of actomyosin fibers in relation to their response to ATP, if this analysis should be technically possible.

When we speak of the shortening and lengthening of actomyosin as a basis of the contraction and relaxation of the muscle, we merely imply at this stage a "vectoral" relationship in their changes of length. The exact manner in which the physical changes in the contractile elements are translated into those in the muscle as a whole is not known. It is, perhaps, in this mechanism that the explanation for the latency lies, and, particularly, that of latency-relaxation (Sandow, 1944) which is found to be an inherent characteristic of muscular contraction. So far, no latency-relaxation has been observed in actomyosin. Possibly this is a technical failure, but more probably it denotes a genuine absence of such a response in the protein fiber. At any rate, the latency-relaxation in an intact muscle, while, in all probability, a genuine physiological process, is hardly to be confused with the ATP-relaxation in actomyosin sol or gel. The difference between the two processes in magnitude and in time-relation to the complete cycle is very considerable. Furthermore, the ATP-relaxation of actomyosin is dependent, to a very large extent, on the enzymic breakdown of ATP. It seems highly improbable that ATP-splitting also occurs during the latency-relaxation period in the muscular contraction.

On the "chemical" side of the mechano-chemistry of muscular activity, the importance of electrolytes in affecting the colloidal behavior of the contractile proteins cannot be overemphasized. Our present knowledge of these electrolytes with respect to their concentration and location in the muscle fibrils, and their movement during contraction, is largely incomplete. The scheme outlined above is intended to stimulate further investigation along this line. In conclusion, it may be emphasized that the far-flung importance of a satisfactory mechano-chemical theory of muscular activity lies beyond and above the understanding of the nature of one specialized physiological process. It goes far to elucidate one of the most fundamental problems in biology, namely, the causal relationship between biological form and function.

LITERATURE CITED

Astbury, W. T., & S. Dickinson
1940. Proc. Roy. Soc. London B **129**: 307.

Bailey, K.
1942. Biochem. J. **36**: 121.

Banga, I., & A. Szent-Györgyi
1942. Stud. Inst. Med. Chem., Univ. of Szeged (Basel-New York) **5**: 25.

Bernal, J. D.
1938. Perspective in Biochemistry: 50. Cambridge University Press.

Brücke, E.
1858. Sitzungsber. k. Akad. Wissensch., Math.-naturw. Cl. Wien **15**: 69.

Caspersson, T., & B. Thorell
1942. Acta Physiol. Scand. **4**: 97.

Dainty, M., A. Kleinzeller, A. S. C. Lawrence, M. Miall, J. Needham, D. M. Needham, & Shih-Chang Shen
1944. J. Gen. Physiol. **27**: 355.

Dubuisson, M.
1942. Arch. Int. Physiol. **52**: 439.

Edsall, J. T.
1942. Adv. Coll. Sci. **2**: 269.

Edsall, J. T., & J. W. Mehl
1940. J. Biol. Chem. **133**: 409.

Engelhardt, W. A.
1942. Yale J. Biol. & Med. **15**: 21.

Engelhardt, W. A., & M. N. Ljubimova
1939. Nature **144**: 668.

Engelhardt, W. A., M. N. Ljubimova, & R. A. Meitina
1941. Compt. rend. Acad. Sci. U. R. S. S. **30**: 644.

Kalckar, H. M.
1944. J. Biol. Chem. **153**: 355.

Lawrence, A. S. C., J. Needham, & Shih-Chang Shen
1944. J. Gen. Physiol. **27**: 201.

von Muralt, A., & J. T. Edsall
1930. J. Biol. Chem. **89**: 315.

Needham, D. M.
1942. Biochem. J. **36**: 113.

Robinson, J. R.
1939. Proc. Roy. Soc. London A **170**: 519.
Sandow, A.
1944. J. Cell. & Comp. Physiol. **24**: 221.
Straub, F. B.
1942. Stud. Inst. Med. Chem., Univ. Szeged (Basel-New York) **2**: 3.
Szent-Györgyi, A.
1946. J. Coll Sci. **1**: 1.
Weber, H. H.
1934. Arch. Ges. Physiol. **235**: 205.

LATENCY RELAXATION AND A THEORY OF MUSCULAR MECHANO-CHEMICAL COUPLING

By Alexander Sandow

Washington Square College of Arts and Science, New York University, New York

The purpose of this paper, broadly speaking, is two-fold. First, there will be discussed certain aspects of the latency relaxation (LR), the minute pre-contractile elongation of a stimulated muscle that occurs in the latter half of the latent period, a phenomenon first observed by Rauh (1922). Evidence will be presented to show that the latency relaxation is a real lengthening of the muscle's contractile material that intervenes between the instant of stimulation and the usual shortening or tension development in a contraction. Further, certain studies of the LR will be considered that suggest it is an indication of a process by which myosin is activated for contraction. The second part of this paper will be concerned with a kinetic theory of mechano-chemical coupling which, in part, is based on the particular significance assigned to the LR.

The LR was discovered by Rauh in the course of some experiments designed to settle the then almost 70-year-old question as to whether the latent period exists at all. To forestall criticism that the demonstration of a mechanical latency merely represented an inertial delay in the response of the registering device, Rauh used an extremely sensitive mechano-optical isometric lever system and photographically recorded the mechanical change during only the first few milliseconds of the twitch response of initially tensed frog sartorii and gastrocnemii. To his astonishment, his records not only exhibited a true delay in the onset of tension development, but they also included a transient negative tension change (the LR) during the latter half of the latent period.

METHOD

Modern electronic methods permit us to achieve vast improvement over Rauh's method. In the work to be reported here, use is made of the piezo-electric cathode-ray oscillographic technique described fully elsewhere (Sandow, 1944). An outline of the method will be given here. A muscle, generally the frog sartorius, is supported inside a moist chamber and, under a few grams resting tension, is connected to the stylus of the piezo-electric pickup, which is a common type of

crystal unit used in phonographs. The pickup is connected to an amplifier and this, in turn, to the vertical plates of a cathode-ray oscillograph. When the muscle is stimulated by a thyratron-tripped condenser discharge produced by an electronic stimulator, a series of tension variations are induced in the muscle which are transformed by the pickup into a corresponding electric pulse that ultimately appears as a set of vertical deflections of the cathode-ray beam. The deflections are spread out on an appropriately timed single sweep of the oscillograph, which is tripped by the stimulating shock. The time-calibration of the sweep is provided by a 10,000 cycles per second sinusoidal voltage supplied by an oscillator, and it is simultaneously impressed on each record, along with the muscle tension change.

In this paper, we shall deal only with the early part of the response, *viz.*, that concerned with the latency relaxation. PLATE 8a illustrates a typical record. The beginning of the sweep at the extreme left is coincident with the instant of stimulation. After a mechanically quiescent period of about 1.4 milliseconds, the sigmoid-shaped latency relaxation is traced out in about 1.8 milliseconds, and this is then rather abruptly terminated as the myogram proceeds very rapidly into the period of tension development. The inset of this figure gives the symbols of the variables that are generally measured. The L's represent latency time intervals all measured from the instant of stimulation: L_R, to the first sign of relaxation; L_0, to the point of inflection of the sigmoid relaxation curve; L, to the point of deepest relaxation; and L_1 to the instant at which the rising tension crosses the base line. The R's measure the magnitude of the relaxation at the instants indicated. In the following discussion, we shall deal almost only with L_R, which measures the relaxation-latency; L, which measures the tension-latency; and R, which measures the depth of LR.

Representative durations of L_R and L have been indicated above. In general, these latencies for a muscle under constant conditions are repeatable within a standard error of $\pm$ 0.03 milliseconds. In the record before us, R corresponds to a negative tension change of about 30 mg. with an error of about $\pm$ 1.0 mg., or an increase in muscle length of about 0.1μ with an error of approximately $\pm$ 0.003μ. It is pertinent to note that the electronic method for recording the LR is so powerful, in converting these minute latency mechanical changes into relatively huge cathode-ray screen deflections, for the simple reasons that, first, the crystal pickup is a very sensitive mechano-electrical transducer, and the 30 mg. tension decrease thus causes a crystal output of about 6 mv.; and, second, such voltages are easily amplified.

In effect, the apparatus acts as an electronic lever, and when the voltage amplification is, say, 5,000 X, as it is generally in our work, the length magnification of muscle changes into screen deflections is about 500,000 X.

Finally, on the original record, an indication, obtained by ordinary optical myography, of the initial tension and peak isometric tension of the muscle is also included in the positions of the bands of light in the upper right corner.

THE CAUSATION OF THE LATENCY RELAXATION

A fundamental question that must now be discussed is the possibility that the LR is an artifact which, in some manner, is a consequence of the muscle's reaction to the contraction set up by stimulation. Schoepfle and Gilson (1945), and Gilson, Schoepfle, and Walker (1947), co-authors in this publication, have raised this question, and have decided that the LR is an artifact, that it may be "interpreted as a phenomenon associated with a compression wave in a fluid system enclosed by an elastic capsule. Compression wave conduction is apparent on comparing local and conducted response records of transverse muscle deflections. A comparable physical model consisting of a water-filled rubber cylinder exhibited an initial deflection in a direction opposite to that of an externally applied force. Hence it is conceivable that the latency relaxation involves a deflection in a direction opposite to that of the internally applied force." (See PLATE 8b.)

Furthermore, they point out that their results "are consistent with those of Ramsey and Street (1940) to the effect that the elastic force of the stretched resting single fiber is associated only with the sheath surrounding the fiber substance. Thus in an isolated resting muscle there is evidently no basic tone and no evidence that active processes diminish such a tone to produce the negative variation."

I believe that these reasons, formidable as they seem, are not sufficient to support the conclusion that the LR is an artifact of mechanical wave transmission. Furthermore, evidence will be presented that permits drawing the conclusion that the LR represents a true physiological elongation precedent to the development of positive tension. In the ensuing discussion, the opposing conceptions of the significance of the LR will be referred to as the artifact interpretation and the physiological interpretation.

No attempt has been made to duplicate Schoepfle and Gilson's experiments with the water-filled rubber tube model of a muscle fiber.

Whatever the results of such experiments, we may, nevertheless, question the validity of the view that the behavior of the model truly represents that of a responding muscle fiber. Quite apart from the obvious general structural differences between the two systems, there is reason to doubt, in particular, the correspondence that has been drawn between the intrasarcolemmar material of a fiber and the water of the model. It is true that this identification is in accord with the conclusion of Ramsey and Street (1940) that the contents of the resting muscle fiber do not exhibit tension. However, Ramsey and Street's result has been questioned by Sichel (1941), who demonstrated that about one-half of the tension of a stretched resting muscle fiber must be attributed to the substance within the sarcolemma.

Moreover, to return to the general procedure of Schoepfle and Gilson, it seems more relevant to inquire as to the behavior of an actual muscle when treated similarly to the model of the fiber. Dr. Gilson, in a personal communication, has informed me that the muscle, unlike the model, does not indicate an initial mechanical change at one end opposite to that externally impressed upon it at the other. I have confirmed this observation in experiments in which a frog sartorius was set up under an initial tension of about 5 gm., so that its pelvic (lower) end was firmly attached to the spring arm of a Harvard 10 mm. signal magnet and its tibial end directly connected to the stylus of the crystal pickup. The signal magnet, in separate tests, was actuated by currents due to application of 4 5 to 0.5 volts, thus varying the rate at which the pelvic end of the muscle was jerked downward. Some results of such experiments are presented in PLATE 9. In each of these records, the sweep, modulated by a 10,000 cycles per second timing-wave, begins at the left at the instant of closing the switch of the signal magnet, and upward deflection corresponds to downward pull by the magnet. Records *A* through *E* represent the responses of the muscle at progressively slower rates of pull, while record *F* was made with the muscle removed and the signal magnet arm directly connected to the pickup stylus. Since the records were all made at the same voltage amplification used in recording the LR response given in FIGURE 1, it is noteworthy that the rate of pull by the muscle on the stylus, especially as in records *A* and *B*, is quite comparable to that introduced by the muscle's own contraction as indicated in the rising phase of the myogram of FIGURE 1. Thus, the effect of an externally impressed mechanical disturbance is being tested under conditions like those that occur in a normal contraction.

Reference to the muscle records now shows that, with the possible

exception of *C, D,* and *E,* the initial response to a jerk at the lower end is a deflection in the same direction at the upper end, which begins after a delay of the order of 0.5 ms., due to the time for transmission of the mechanical wave up the length of the muscle.* Very close examination of records *C, D,* and *E,* however, shows the presence of an extremely small downward deflection preceding the steeply rising upward change. Actually, this occurs in only about 10% of the tests and, thus, where it does exist, it can be attributed to a random irregularity of the base. Alternatively, if these initial downward movements are valid indications of a muscle behavior qualitatively like that of the Schoepfle and Gilson model, they are so minute and so irregular, in comparison with the large uniformly obtained LR's observed in stimulated muscle, that they could not be accepted as evidence supporting the artifact interpretation of the LR.

Thus, it is our conclusion that, contrary to the expectation in terms of the artifact interpretation, the introduction of a mechanical wave initially oriented in a certain direction at one end of a muscle does not lead to an oppositely directed change at the other end.

I shall now consider the experiments of Schoepfle and Gilson involving transverse recording of the LR, limiting the discussion mainly to the crux of the matter, namely, the nature of the response to direct stimulation at the point where the stimulus first excites the muscle, *i.e.,* at the cathode of the stimulating electrodes. In this type of experiment, the muscle, a bullfrog sartorius in the work of Schoepfle and Gilson, was laid on a plate with its ends anchored, so that it was under about 20 gm. axial tension. The pickup stylus was pressed lightly (this tension was not measured) across the long axis of the muscle. If, now, the stylus is made cathode, the latency response is as shown in *B* of PLATE 8b. Part C (PLATE 8b) represents the latency response when the cathode was at a point 2.5 cm. from the stylus. There are obviously great differences between these two records, but no attempt will be made here to discuss all of them. Of critical importance, however, is the fact that, even when the stylus is cathode, a small but definite LR precedes tension onset. Since this LR is recorded at the point of stimulation where, at the very first moment of the mechanical response, transmitted wave effects are not present, the initial elongation in the muscle represented by the recorded LR change must have

* This time interval has been corrected for the delay of the signal magnet in beginning its response, which, as indicated in record F of PLATE 9, is about 0.1 ms. Since the length of the muscle in this experiment was 4.0 cm., the velocity of propagation of the mechanical wave was 8 0 cm /ms. This velocity is much higher than that reported for European frogs by **Fischer** (1926) and by **Buchthal & Knappeis** (1943). Further discussion of this difference will be given elsewhere.

occurred prior to, and hence independently of, the tension wave. Thus, a proper evaluation of Schoepfle and Gilson's results, obtained from both model and muscle experiments, does not support their artifact interpretation of the LR.

I will now discuss certain experiments, performed in my own laboratory, that give quite positive evidence that the LR actually precedes the onset of tension. The first series of experiments involves transverse recording, somewhat as in Schoepfle and Gilson's. FIGURE 1 presents the way the muscle, the frog sartorius or gastrocnemius, is set up for transverse recording. The muscle is slung in air between two pillars of the support, *S*, so as to be under 3 to 5 gm. axial tension. The pickup-stylus, *s*, which is at a right angle to the muscle's long axis may either press down on the muscle from above, as at *A*, or pull up against the muscle from below as at *B*. In either case, the small mirror, *M*, stuck on to the stylus' end, permits determination of the transverse tension applied by the stylus to the muscle, by means of a

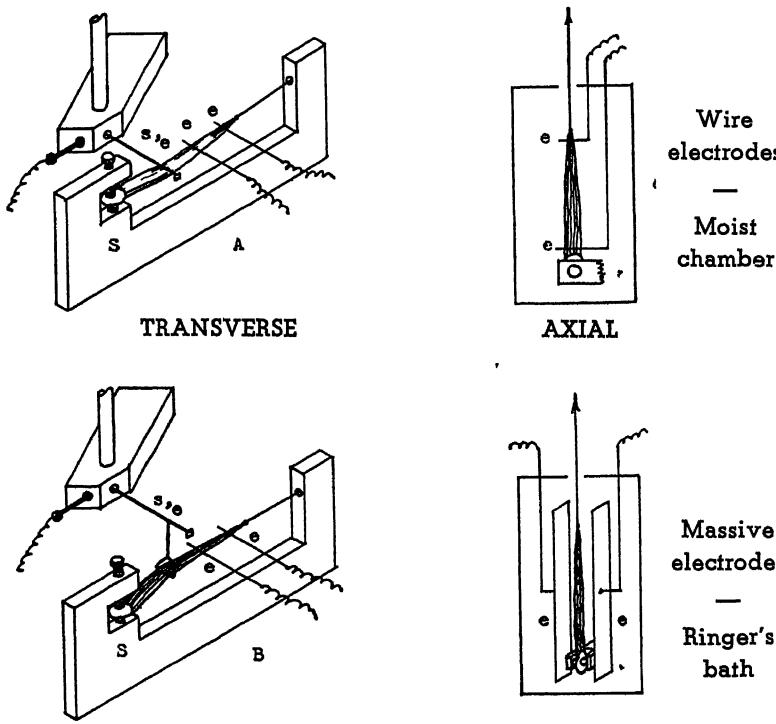


FIGURE 1. Transverse and axial arrangements for recording latency mechanical responses. S, muscle support; s, pickup stylus; e, electrode.

conventional optical lever system. Furthermore, this optical lever system is also the means by which may be recorded the thickening at peak of contraction of the muscle segment in contact with the stylus.

The stylus may be made a stimulating electrode, by connecting a wire from the stimulator to the metal screw which serves to fix the stylus in its chuck. Other, Ag-Ag Cl, electrodes, *e*, are available at other points on the muscle, so that a variety of electrodal configurations for actual stimulation may be used.

PLATE 8c presents a typical transverse LR record obtained from a frog sartorius under 5 gm. transverse tension at the stylus, which acts as cathode. The LR deflection is now directed upwards, simply because the various tensions under transverse recording are oppositely directed to those present in the axial technique. The general shape of the LR, however, is quite the same as that obtained axially. Greatly significant, here, is the large deflection obtained at the site of stimulation. Indeed, it is much larger than that recorded when the cathode is 1 cm. distant. This point, and some others, are demonstrated in FIGURE 2.

The graphs of the left part of this figure present the results of an experiment in which the LR of the sartorius was recorded as a function of transverse tension at the stylus; for each tension, a record was first obtained with the stylus acting as cathode, and then with the cathode 1 cm. distant. I wish to mention only two conclusions of this experiment: (1) The value of R varies with the transverse tension, and (2) R is always larger when recorded at the cathode. The second point is especially interesting, for it is diametrically opposed to the finding of Schoepfle and Gilson for the bullfrog sartorius.

The right part of FIGURE 2 presents the results of a somewhat similar experiment with the frog gastrocnemius. Here again, we see the great variation in R with transverse tension. However, up to about 1.5 gm. tension, R is greater when the cathode is distant from the stylus, in accordance with Schoepfle and Gilson. Beyond this tension, the reverse is true.

Thus, the variation in R with stylus-cathode separation is a function of the kind of muscle and of the transverse tension. Although I have not worked with the bullfrog sartorius, I would hazard the guess that, if the transverse tension were high enough, R, as in the gastrocnemius experiment, would be greater with the cathode at stylus than with the cathode some distance away. The various L's also vary with these conditions, but no attempt will be made here to explain these differences. The most important result of these experiments, how-

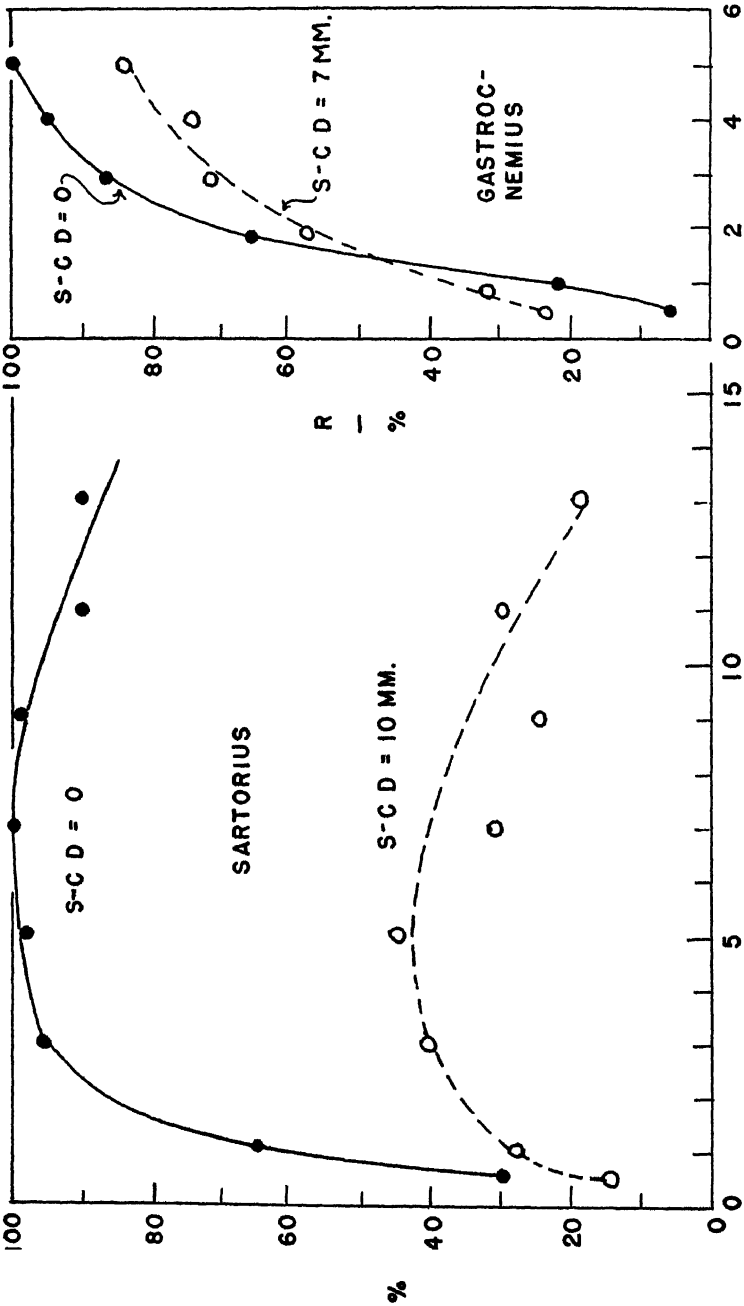


FIGURE 2 Effect of transverse tension on transversely recorded latency responses of frog sartorius and gastrocnemius muscles. *S-C D* refers to the distance between pickup stylus and stimulating cathode. *S-C D* = 0 means that the stylus also acts as cathode. *R* represents the depth of the LR.

ever, is that, regardless of which muscle is used, and provided the transverse tension is not zero, an LR will always be recorded when the stylus is cathode. This crucial result has been further tested on the frog sartorius, in experiments in which special precautions were taken to sharply localize the actual cathodal region on the muscle by using as stylus-cathode a fine piano wire (0.2 mm. thick) and drying up all the electrolyte at this cathode, except that needed to ensure electrical conduction to the muscle. Despite these conditions, the first recorded mechanical change was still the usual initial relaxation. Hence, we must conclude that the initial relaxation is not an artifact due to the transmitted effect of a prior contraction at some distant point of the muscle.

Other experimental approaches to the point at issue are afforded by certain applications of the axial recording technique, as diagrammed in FIGURE 1, in which a sartorius muscle is supported vertically in a chamber with its upper end connected to the pickup stylus by means of a short length of piano wire or fine metal chain. In the case of wire electrode stimulation, the cathode is placed at the very upper (tibial) tip of the muscle. The cathode is thus localized at the extreme end of the muscle and, when excitation occurs, the mechanical response of the muscle is directly transmitted to the pickup, without having to pass longitudinally through any part of the muscle. Hence, wave transmission, in the sense of Schoepfle and Gilson, does not occur. Nevertheless, a quite normal LR is recorded.

The massive electrode technique, used by Brown in 1941 (though first introduced by Brown and Sichel (1936), in somewhat different form), also illustrated in FIGURE 1, involves the use of Ag-AgCl bands, 8 x 50 mm., which are placed to flank the muscle symmetrically on either side, with electrodes and muscle immersed in a Ringer's solution bath to permit electrical conduction of the shock through the muscle. According to Brown, stimulation with such electrodes simultaneously throws into activity all points along the entire length of the muscle. Buchthal and Knappeis (1943) doubt this interpretation, but their argument is not convincing, despite the reference to the difficulty of stimulating with shocks of which the electric field is transverse to the axes of the fibers. For it is doubtful that the use of massive electrodes introduces a strictly transverse electric field. More likely, there are small deviations from the transverse condition along the whole muscle length, and the resulting longitudinal components scattered over the entire muscle are the actual stimulating currents that throw the full length of the muscle into simultaneous response. Thus,

under such conditions, there can be no conducted mechanical wave and, if the artifact interpretation of the LR be correct, no initial lengthening should be observed. Actually, a very deep LR is obtained with a value of R three to four times that observed in the same muscle excited by means of wire electrodes placed near the muscle's ends. The large increase in R is explained by the synchronous activation of the whole muscle length. Thus, under conditions where no compressional wave exists along the muscle length, or where it must certainly be reduced to a minimum, the LR is still present and, moreover, to a larger degree than with other kinds of response.

The above experimental evidence, therefore, does not support the artifact interpretation of the LR. This interpretation contends that the immediate mechanical response of a muscle to stimulation is a contraction at the point of excitation, and that the compressional wave thus set up is first indicated, at a point distant from the site of excitation, as an oppositely directed change, *i.e.*, a relaxation, the LR. However, our experiments demonstrate that, when a mechanical deformation is externally impressed upon one end of a muscle, the other end—contrary to the artifact interpretation- is deformed essentially in the same sense. More important, however, is the behavior of physiologically activated muscle which is studied in such manner that the effects of a mechanically propagated contraction wave cannot possibly play any part in determining at least the very initial portion of the tension variations which make up the latency behavior (*i.e.*: transverse recording with the stylus acting as cathode; axial recording with the cathode at the extreme end of the muscle connected to the stylus; and axial recording with massive electrodes for stimulation). Under these conditions, the initial tension change is not contraction, as is required by the artifact interpretation, but a relaxation, the LR. In view of this mass of evidence, we must conclude that Schoepfle and Gilson's artifact interpretation of the LR is not correct, and that the cause of the LR must be sought in some mechanism other than the transmitted effects of the compressional wave set up by the contraction of a muscle.

The proposal of Schoepfle and Gilson is, to date, the only formal attempt to prove that the LR is an artifact based on a mechanism which can be subjected to rigorous test. The disproof of this mechanism as a basis for the LR does not necessarily mean that the physiological interpretation is thus assured unquestioned validity. Other possibilities for an artifact have been informally suggested: (1) That the LR is due to "slippage" of fibers, and (2) that it is due to the "flip" of the muscle. Both these mechanisms are rather vague and thus difficult to

test experimentally. It would seem, however, that "slippage" and "flip" might vary somehow in the separate contractions of a given muscle in which its initial orientation was changed, by varying, either separately or together, the twist of the muscle about its longitudinal axis, and the sheer of the fibers in the muscle, by rotating the attached pubic symphysis in the clamp which fixates the pelvic end of the muscle. It is significant, therefore, that, despite the introduction of such alterations, the LR regularly appears.

In view of all the above evidence, it seems reasonable to conclude that the LR is not an artifact, but that it represents the first physiological mechanical change in an excited muscle, an initial slight relaxation that just precedes the onset of positive tension development.

SOURCE OF THE LATENCY RELAXATION

Since it is clear that we may deal with the LR as a true precontractile elongation of stimulated muscle, it is pertinent to inquire whether the LR has its source in the response of the excitatory or the mechanical system of muscle. The special pertinence of this question lies in the possibility that at least some phases of the excitatory process at the cathode and the very early transmission of the action potential may be quite synchronous with the occurrence of the recorded LR. It is possible, however, to rule out the excitation mechanisms as the source of the LR. First, because, in an excited nerve trunk (the frog sciatic) in which excitation and conduction are presumably similar to these processes in muscle, nothing like an LR appears when it is connected to the pickup stylus in the same general way that muscle is tested. Furthermore, in general, there is no parallelism in the behavior of the parameters of the action potential and of the LR when the muscle state is altered by such factors as activity or temperature.

An illustrative test involving the effect of activity is outlined in the following.* Sartorius muscles were subjected to a series of 2 sec. tetani, at an average rate of one tetanus each 7 sec. In each period intervening between two successive tetani, some muscles were tested for their LR, and others had their action potentials (diphasic) recorded. Some of the results of such experiments are presented in the graphs of FIGURE 3. In this figure are plotted the relative average values (12 experiments) of L and the relative average values (2 experiments) of t, which represent the time from the instant of stimulation to the beginning of the recorded action potential, which is mostly conduction

* The experiments of this test, and of the studies with pH and temperature discussed later, were performed with the collaboration of Mr. A. G. Karczmar, and were supported, in part, by a grant from the Penrose Fund of the American Philosophical Society.

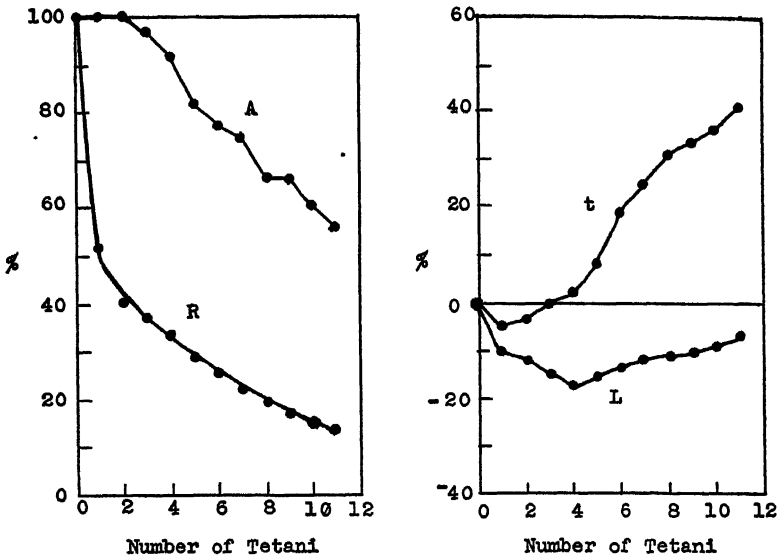


FIGURE 3. Effects of a series of 2 sec tetani at 7 sec. intervals on certain parameters of the latency response and the action potential of a frog sartorius. Latency variables as defined in PLATE 8a. t, time from instant of stimulation to foot of diphasic action potential; A, height of first limb of action potential.

time. It is evident that the plotted time parameters of the two phenomena follow different variations in response to the successive periods of tetanus. Although both L and *t* first fall and then rise with increasing activity, L reaches a maximum decrease of 17 per cent at the fourth tetanus, but *t* has a greatest decrease of only 6 per cent which is reached after but one tetanus. Furthermore, during the latter tetani of the set of 12 studied, L, though rising, does not reach its original value, while *t* increases to a final value about 41 per cent greater than it was originally. Nor does L_R (not plotted) show any better correlation with *t*, for this LR parameter is substantially constant during an activity series such as studied here. FIGURE 3 also presents the changes in R and A (the value of the magnitude of the peak of the first limb of the action potential). It is noteworthy that, although the first two tetani cause no significant change in A, they cause a sharp fall in R. Beginning with the effect of the third tetanus, R and A both fall, though A is always much greater than R. Thus, neither the time nor magnitude parameters of the LR and the action potential vary in parallel in response to activity.* Further evidence of this sort

* The validity of this conclusion may be questioned, in view of the fact that, on account of experimental difficulties, the action potentials were not recorded at the same point of the muscle at which the latency mechanical events were registered. In refutation of this criticism may be mentioned the result discussed in

will be presented later, in connection with our discussion of the effects of temperature on the LR. Thus, in general, the LR and action potential variables seem to alter independently of each other. We must, therefore, conclude that the LR does not have its source in the mechanisms responsible for the action potential, and hence we are led to the view that the precontractile elongation is associated with some feature of the behavior of the contractile material. Other evidence, to be presented later, will confirm this view, and will show, moreover, that it is the myosin, in particular, that is responsible for the LR.

Before entering this phase of the discussion, however, it is pertinent at this moment to take up the previously mentioned point concerned with Ramsey and Street's (1940) result that the resting tone of a single muscle fiber is localized in the sarcolemma and not in the contractile substance. How can we accept our view that the LR (which involves a reduction in tension from that found at rest) is associated with the contractile matter, if this material has no basic tone? In answer to this question, it must be noted, in the first place, that the LR tension change in a *single fiber* is extremely small. If the LR tension decrease of an entire sartorius is, say, 30 mg., and we assume, as a reasonable figure, that there are about 1500 fibers in the muscle, then the LR decrease in tension per fiber is only 0.02 mg. Since Ramsey and Street's error in measuring tensions of their fibers was certainly of the order of one milligram, it is clear that they could not have detected a small residual tension of the contractile material that would be ample to account for the decrease in tone of the contractile material connected with the LR. Furthermore (disregarding Sichel's (1942) finding to the contrary), let us assume that there is absolutely no tone in the contractile substance of a resting muscle which, as a whole, is under tension. We all agree that, during a response, the contractile material somehow develops tension that is ultimately transmitted to our registering devices. It is, therefore, not unreasonable to suppose that, during the LR, some forcible lengthening change occurs in the contractile substance that is likewise transmitted to our registering apparatus. Hence, we see that it is possible to remove another support to the view that the LR is an artifact, and we are thus further assured that the LR indicates some real underlying process in the fundamental mechanism involved in activating the contractile machine to develop positive tension.

an earlier paper (**Sandow**, 1945b) that the activity-induced changes which alter the latency behavior extend throughout the whole of the muscle. Thus, wherever the action potentials are recorded, the excitatory mechanisms in which they have their source would be subjected to the same alterations in the muscle's internal milieu that modify the system responsible for the LR.

EFFECT OF pH ON THE LATENCY RELAXATION

I will now pass to a discussion of certain experiments that seem to shed some light on the nature of this process, first dealing with the effect of pH on the events of the latent period. Changes in muscle hydrogen-ion concentration have been affected in two ways: indirectly, by the muscle's own activity, and directly by soaking the muscles in solutions of varying acidity. The first approach has been treated fully in a recent publication (Sandow, 1945b), hence it is necessary here only to state the quite clear conclusions of this work. As a consequence of progressive activity of the muscle (a series of twitches at 3 sec. intervals): (1) R progressively falls; (2) LR remains constant, except in extreme fatigue, when it rises; and (3) the tension-latency, as represented not only by L and L_1, but also L_0, first rises, then falls, and then rises again. In general, these changes are reversed by rest after the activity. No explanation for the behavior of R and L_R is at present available. But the tension-latency variations may be quite definitely correlated with the activity-induced pH changes, *i.e.*, the higher the pH, the shorter the tension-latency.

Comparable results are obtained by directly modifying the pH, through soaking the muscle in Ringer's solutions of varying pH determined by proper adjustments of a bicarbonate buffer system. The results of such an experiment are presented in FIGURE 4. The explanation is not yet at hand for the sharp drop in R at pH's below 7.5 (The actual experiment behavior above this pH is, at present, unsettled.) LR is fairly constant above pH 7.0; as for its rise at lower pH, this may be compared with a similar rise in L_R in a very fatigued muscle where it is known that the pH falls due to lactic acid accumulation. I wish to call attention, especially, to the general fall in the tension-latency over the entire range studied from pH 6.0 to 8.5, and to the gradual decrease in the time interval L_1-L_R from about pH 6.5 to 8.5. Here, again, the higher the pH, the shorter is the latency for tension output. But even more suggestive is the fact that the interval L_1-L_R at the higher pH's falls, thus indicating that, in this range at least, a process localized in time between the beginning of the LR and the time when actual tension is developing is more rapid in rate, the higher the pH. Before passing to an interpretation of these tension-latency alterations, it may be noted that they are in general agreement with similarly oriented results obtained by Ritchie (1933) for the usual latent period in heart muscle.

In signifying the meaning of the pH results, it seems very pertinent

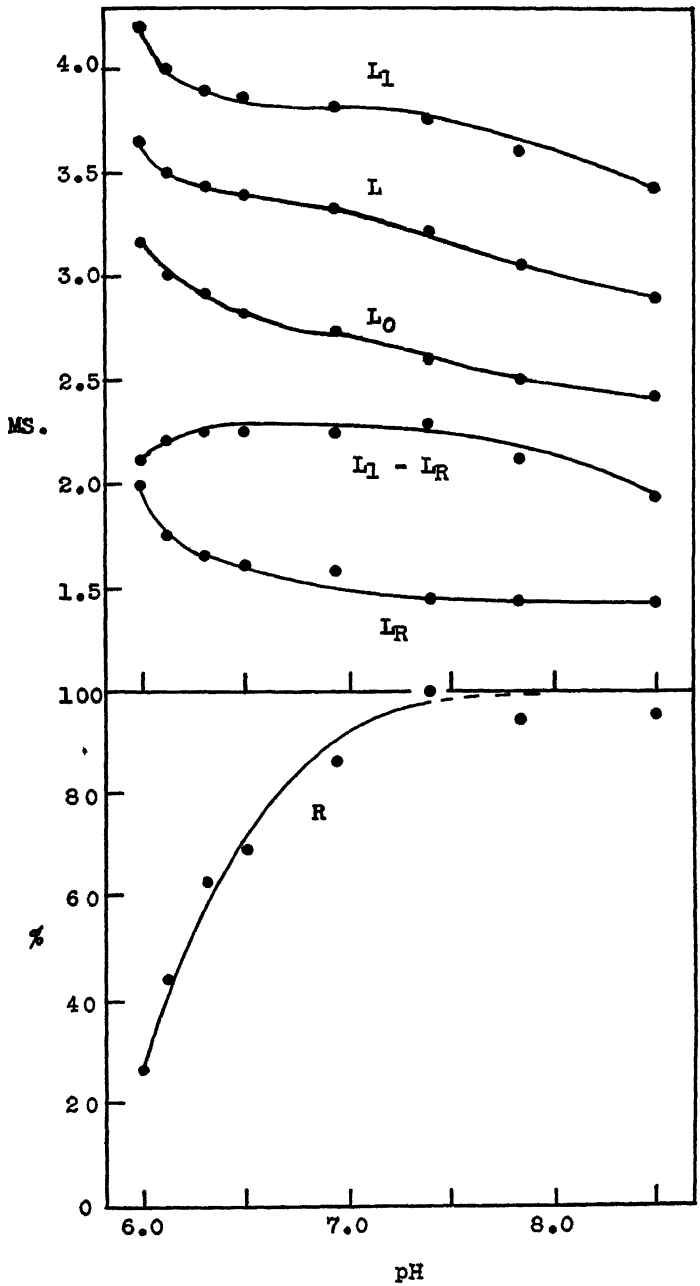


FIGURE 4 Effect of pH on latency variables (The dashed portion of the line for R indicates that the behavior of R above pH 7.5 is, at present, uncertain)

that the rate of hydrolysis of ATP by myosin-ATPase, is, at least up to pH 9.0, more rapid the higher the pH. This behavior has been demonstrated by Bailey (1942) and DuBois and Potter (1943), and by Singher and Meister (1945).

The fact that the rate of tension onset, as measured by the tension-latency, and the rate of ATP hydrolysis by myosin-ATPase vary in the same direction with pH suggests that, in the stimulated muscle, tension development is determined by the breakdown of ATP. In already published work (1945b), I have discussed at some length the reasons for believing that the role of ATP is quite unique in accounting for the effect of pH on the tension latency. The view has therefore been adopted that, when a muscle is stimulated, something occurs that permits ATP and myosin to interact during at least the latent period. In this interaction, it is supposed that the first energy-rich phosphate bond of ATP is split, thus yielding ADP and phosphate, and that the energy liberated is immediately taken up by the myosin, which, thus energized, then contracts. Furthermore, the fact that, in the higher pH range, the interval L_1-L_R decreases with pH, is taken to mean that the ATP-myosin interaction does not begin until the LR has begun. Thus, the interval L_R is supposed to be a space of time in which processes preparatory to the interaction of myosin-ATPase and ATP take place. The above assumptions are obviously in general agreement with the well-established view that ATP is the immediate source of energy for myosin activity, and that, as established by Engelhardt and other workers, myosin, or an enzyme closely associated with it, is ATPase.

EFFECT OF TEMPERATURE ON LATENCY RELAXATION

If the ATP hydrolysis that is supposed to take place in the latent period (specially beginning with the onset of the LR) is catalyzed by myosin-ATPase, then it is important to search for independent proof of the involvement of myosin in this process Evidence has already been adduced suggesting that the contractile material, myosin, is somehow involved in the causation of the LR. However, certain experiments on the effects of temperature on R more definitely specify this role of myosin.

FIGURE 5 presents the results of a typical experiment to test the effect of temperature on the latency parameters. Although this set of graphs includes data on the behavior of the time variables, the present discussion will be concerned only with R. As the temperature is

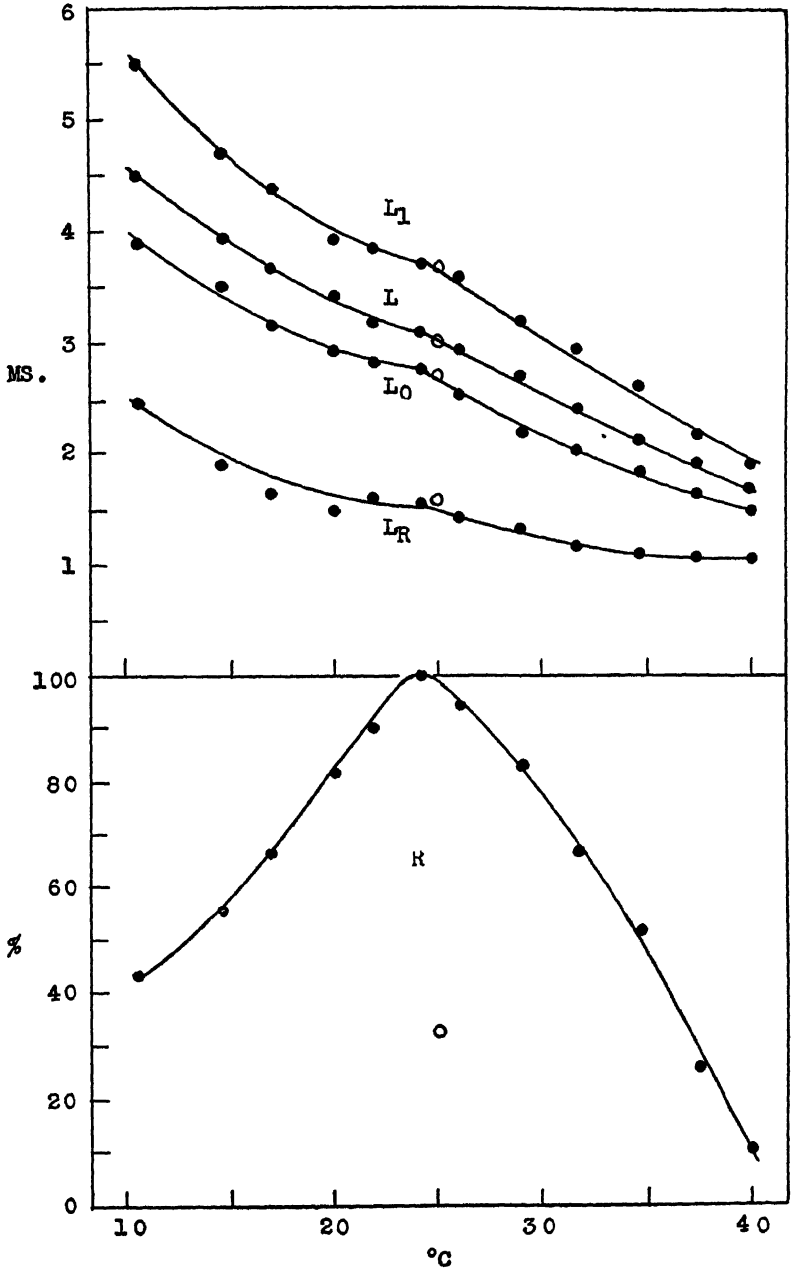


FIGURE 5 Effect of temperature on latency variables.

varied from 10° to 40° C., R increases to a maximum at about 24° and then falls at higher temperatures until, at 40°, it is only a few per cent of the maximal value.* If the muscle is kept for a minute or so at 40° and then quickly returned to the optimal temperature, 24°, the magnitude of R (as indicated in FIGURE 5 by the circle) is greater than it was at 40°, but not as great as it was at 24°, before submission to the higher temperature. Thus, the inactivation of R at high temperature can be partially reversed. The detailed study of inactivation of R by high temperatures leads to the following conclusions:

(1) The irreversible inactivation of R begins quite sharply at a critical temperature of about 37.5°, and the rate of irreversible inactivation increases very rapidly within the range of the next few degrees. The temperature coefficient of this process at about 38°–39°, expressed as a Q_{10} value, is of the order of several thousand.

(2) Another critical temperature exists at about 41°, for, at or above this temperature, completely irreversible inactivation of R occurs very quickly, *i.e.*, in a few minutes.

The general variation of R with temperature is similar to that characteristic of temperature effects on enzyme behavior. Therefore, it is inferred that the LR is an expression of the action of an enzyme. Furthermore, the two critical temperatures for the inactivation of R, one at 37.5°, the other at about 41°, are, respectively, essentially the same as those at which Mirsky (1937) has demonstrated the occurrence of primary and secondary heat denaturation of extracted frog myosin, and the high Q_{10} for the LR inactivation at 38–39° C. is comparable to that of the order of one thousand found by Mirsky at a similar temperature range for the primary denaturation process. The thermal features of the inactivation of the LR and the denaturation of myosin are so similar that we therefore conclude the LR is determined by some action of myosin. Since these generalizations concerning the significance of the temperature variations in R indicate that, on the one hand, the LR represents an enzyme action, and, on the other, a property of myosin, it is reasonable to infer that the LR is an indication of myosin acting in close association with some enzyme. This view is given more definite form by the inference, derived by analysis of the effect of pH, that the hydrolysis of ATP, the substrate of myosin-

* Some preliminary studies of the variation in the magnitude of the action potential of the frog sartorius, over the same range of temperature, show that this quantity rises in value up to about 20° C., then falls to a minimum at about 24–25°, then rises again to a new maximum around 35°, and finally, at still higher temperatures, falls until, at about 41°, it is practically zero. This behavior is different from that of R. Thus, these results add further support to the previously drawn conclusion that the LR has its source in a different system from that which produces the action potential.

ATPase, is occurring during the LR phase of the latent period. Hence, the general picture obtained from these considerations is that the latent period, or more specifically the LR part of this period, is a time interval during which at least the very initial phase of the coupling between myosin-ATPase and ATP takes place. Furthermore, it is supposed that, as a result of this coupling, ATP is immediately split, and the energy so released is used to activate myosin directly for contraction. This scheme is, thus, a form of activity energization.

A THEORY OF MECHANO-CHEMICAL COUPLING

The question that now arises concerns the detailed nature of the mechano-chemical coupling of myosin-ATPase and ATP, and its relation to the general sequence of mechanical changes in contracting muscle. The following theory, which has appeared in a preliminary form (Sandow, 1945a), is proposed to account for these events.

The central feature of the theory is an assumed cycle of transformations of myosin-ATPase, M, as follows:

Stimulus → Excitatory Process → Myosin-ATPase Activator

$$\begin{aligned} M + A &\xrightarrow{k_1} M_A \\ M_A + ATP &\xrightarrow{k_2} M_A\text{-}ATP \\ M_A\text{-}ATP &\xrightarrow{k_3} M_C + ADP + P \\ M_C + ? &\xrightarrow{k_4} M + ? \end{aligned} \qquad \text{(I)}$$

It is supposed that the essential function of the stimulus is to lead, by means of the excitatory process, to the formation of an activator for the myosin-ATPase system. Little is known of this process in live muscle, although various experiments with myosin extracts (Bailey, 1942; Szent-Györgyi, 1946; Brown, 1946) suggest that probably the activator is Ca^{++} or K^+. At any rate, it is presumed that time is required for the release of the activator before combination with myosin can occur, and it is our hypothesis that the time interval L_R in our records represents this delay.

The first reaction involving myosin-ATPase, M, now takes place in the form of a combination with the activator A, thus forming M_A. The second myosin reaction is the production of an enzyme-substrate complex between M_A and ATP, forming M_A-ATP which we assume to

be the state of myosin responsible for the mechanical change observed in the latency relaxation. In support of this view, or at least of the conception that a myosin-ATP interaction is associated with an elongation of the myosin, we note the following.

Engelhardt (1941, 1942), it will be recalled, in addition to demonstrating that myosin extracts have ATPase activity, also reported that his artificially spun myosin fibers undergo an increase in extensibility specifically in the act of splitting ATP. This strongly suggests that the energy transfer from ATP to myosin in live muscle is a function of the enzymatic activity connected with a relaxation phenomenon. Hence, Engelhardt suggested that the interaction of myosin and ATP in the intact muscle occurred during the post-contractile relaxation. However, he was evidently unaware of the existence of the precontractile latency relaxation, and, if muscular relaxation is an *in vivo* sign of the interaction of myosin and ATP, then this reaction may be going on at the time of the LR. At any rate, in view of the experimental evidence already presented indicating ATP breakdown catalyzed by myosin-ATPase during the latent period, that is the assumption made here.

Our assumption goes further, however, for we suppose that it is the enzyme-substrate phase of the interaction of myosin and ATP that is indicated by the LR. This, of course, is an inference. It is based on the view that Michaelis' conception of enzyme action holds for myosin-ATPase, and that the formation of the intermediary complex follows the activation of the enzyme. But it is not clear, in detail, how our hypothesis may be harmonized with Engelhardt's observation of the increase in extensibility of myosin fibers under ATP action, for these fibers, although splitting ATP, presumably retained their greater extensibility, and certainly did not contract. This limitation in the behavior of Engelhardt's fiber may be due to the operations required to extract myosin and form it into threads. Our assumption is, furthermore, in direct contradiction to the view of the Needham group's work (Dainty *et al*, 1944), which led to the inference that *contracting* myosin is represented by the M_A-ATP combination. This work, however, was performed on myosin *micellae in solution,* and, as the Needham paper states, "It must be remembered that, *in vitro,* the contraction of the myosin particles does no work, and the conditions are therefore very different from those *in vivo.*" In fact, energy from some source is being transformed by the contracting muscle, partly into the heat liberated concurrently with tension rise in the contraction period, and partly into the mechanical potential energy of the ten-

sion itself. It is difficult to reconcile these energy transformations with the Needham view that an enzyme-substrate combination is associated with the contraction period. For, if the ATP, the source of the energy, is still in combination with its enzyme, then the energy of its first phosphate bond is not yet available for release into other parts of the system. Furthermore, the mechanical effect of ATP on Engelhardt's fibers is different from that on Needham's micellae, for the former merely are more extensible, while the latter seem to contract and then relax. Probably, in live muscle, a synthesis of these myosin mechanical changes occurs; *i.e.*, the LR corresponds to the elongation of Engelhardt's fibers, and the subsequent contraction and relaxation are equivalent to the shortening and then lengthening of the Needham micellae. The evidence for the assumption that the LR representing the enzymatic intermediary is thus not unequivocal. Nevertheless, we are adopting it as a working hypothesis. Since the M_A-ATP complex is assumed to be a relaxed form of myosin, it will be symbolized by M_R.

The third reaction in the chain is now assumed to be $M_A\text{-ATP} \rightarrow M_C + ADP + P$, where P is phosphate, ADP adenosinediphosphate, and M_C, myosin under contraction and energized for this activity by the absorption of the energy released by the splitting off of the first phosphate of ATP. Probably M_C is another enzyme-substrate complex, now of M_C and ADP.

The final reaction is $M_C \rightarrow M$ and, mechanically, this brings about the post-contractile relaxation of the myosin. This process may be connected to some reaction involving ADP, but the evidence presented here gives no indication what this may be. See, however, the Introduction (p. 667) for further discussion of this point.

We now have for the cycle of myosin transformations in our scheme, the following reaction chain:

$$M \xrightarrow{k_1} M_A \xrightarrow{k_2} M_R \xrightarrow{k_3} M_C \xrightarrow{k_4} M \qquad (I')$$

It is assumed that, to a first approximation, the individual reactions are all first order, the various k's representing the velocity constants. Let us indicate the concentrations of the various myosins by corresponding small m's, and use a rational system of concentration units, such that, with the muscle at rest, $m = 1$. Since it requires a stimulus to set the reaction chain (I') going, in the absence of a stimulus, *i.e.*, in the muscle at rest, myosin only exists as M. Hence, the initial conditions for the reaction chain are:

$$m = 1,\ m_A = m_R = m_C = 0 \qquad (II)$$

It is interesting to note that the kinetic system represented by the reaction chain (I′) and the initial conditions (II) are formally similar to a kinetic basis for the course of tension output suggested on quite different grounds by Schoepfle and Gilson (1945), and Gilson, Schoepfle, and Walker (1947).

The solutions of the kinetic system (I′) giving the various m's as a function of the time are well known. It is not necessary to list the solutions here; it is sufficient merely to state that, in general, each function for a particular m includes a sum of exponential terms whose exponents involve the k's and whose coefficients are functions of the k's. Thus, in general:

$$m_N = \sum_{i=1}^{N} c_{N_i} e^{-k_i t} \tag{1}$$

where, N represents the number of the product in the chain (I′):

$$m_1 = m,\ m_2 = m_A,\ m_3 = m_R,\ \text{etc.} \ldots$$

and,

$$c_{N1} = \frac{k_1 k_2 \cdot \cdot \cdot \cdot \cdot k_{N-1}}{(k_2 - k_1)(k_3 - k_1) \cdot \cdot \cdot \cdot \cdot (k_N - k_1)},$$

$$c_{N2} = \frac{k_1 k_2 \cdot \cdot \cdot \cdot \cdot k_{N-1}}{(k_1 - k_2)(k_3 - k_2) \cdot \cdot \cdot \cdot \cdot (k_N - k_2)},$$

etc.

The assumption is now made that the tension, T, at any instant, is given by some function of the instantaneous concentrations of M_R, which will introduce a negative tension change, and M_C, which will cause a positive change. The simplest form of this assumption is that, numerically,

$$T = m_C - a m_R \tag{2}$$

The constant, a, may be designated as the tension coefficient of m_R and refers to the view that, while a unit concentration of M_C generates a unit of positive tension, this is not true for the negative tension associated with M_R. Indeed, it will be shown later that a must be much less than unity.

To obtain numerical values of T from EQUATION 2, it is obvious that we must have the k's numerically evaluated, and this should be done, where possible, not arbitrarily in a process of curve-fitting, but on an independent, experimental basis. To this problem, we shall now turn.

There seems, at present, to be no clear experimental basis for evaluating k_1. The reaction, $M \rightarrow M_A$, however, must be a very rapid process, since it is contained probably only in the first millisecond or

so of the latent period. Hence, somewhat arbitrarily, k_1 has been assigned the rather large value of 500 sec.$^{-1}$.

A rather definite basis for determining k_2 may be found if it be assumed that the formation of M_A-ATP, *i.e.*, M_R, be the chemical reaction underlying Brown's alpha process. According to Brown (1941), several lines of work indicate that, very soon after stimulation, *i.e.*, in about 0.1 of the contraction period, this process, which is intimately related to the activation of contraction, reaches a peak and then falls rapidly, exponentially, so that it has practically ended at the peak of contraction. Now, there are several reasons for assuming that the alpha-process may be represented by the LR, and thus according to our scheme by the course of M_R. First, the experimentally determined time of greatest activity connected with the alpha-process is coincident with the development of the LR. Second, Brown (1946) infers that the material of importance in determining the alpha process is a complex of myosin-ATPase, ATP, and Ca^{++}, *i.e.*, essentially the same activated enzyme-substrate combination that we assume for myosin in the LR. And, third, this complex is formed with a reduction in molecular volume (Brown and Lawler, 1946), which, in the muscle, might be accompanied by an LR *elongation* of the myosin fibers, in view of the anisodiametric length changes believed to occur in myosin micellae with changes in volume.* That is, since the protein micellae of muscle are supposed to shorten upon swelling, they would be expected to lengthen when their volume is reduced. Thus, there exists a good basis for evaluating k_2 in terms of the kinetics of the alpha-process. On this basis, k_2 has been set at 200 sec.$^{-1}$.

The value of k_3 has been determined on the following basis. Reference to the sequence of reactions (I') indicates that, if $k_4 = 0$, that is, if the post-contractile relaxation reaction is stopped, all the myosin, as time goes on, will end up in the form of M_C, thus producing a condition for the maximum tension output of which the theoretical system is capable: *i.e.*, the system behaves in a manner corresponding to the production of maximum tension output of a muscle in a tetanus. We now assume that the contribution of m_R may be neglected in the approximate determination of the course of the rise of tetanus tension, this being justified on the basis of preliminary calculations which show that a, the coefficient of m_R, is small (of the order of a few hundredths),

* These considerations, although theoretical, suggest a possible mechanism for the production of the LR which is connected to the physiological behavior of the muscle. The previously mentioned results concerned with temperature similarly point to a physiological basis for the LR. To the extent that such analyses are valid, they help in further strengthening the physiological interpretation of the LR.

and that m_R will itself be rather small in the latter part of the interval during which the tetanus is developing. In addition, to further simplify the procedure, we neglect in the equation for the tetanus the terms having the exponentials $e^{-k_1 t}$ and $e^{-k_2 t}$, since, at the relatively large times that apply to the tetanus-rise, the values of the exponentials are so small. Thus, under these conditions, EQUATION 2 reduces to $T = 1 - [k_1 k_2/(k_1 - k_3)(k_2 - k_3)]e^{-k_3 t}$. Now, by substituting trial values for k_3 in this equation (k_1 and k_2 being known), a value of k_3 is found that gives a theoretical curve showing a reasonable fit to typical experimental tetanus curves. Thus, a rounded average value of k_3 for frog skeletal muscle at about 25° C. has been found to be 45 sec.$^{-1}$, and this is adopted in the present work.

The evaluation of k_4 is carried out by a somewhat similar method. It is now supposed that the system corresponding to (I′) has been placed into a state of full tetanus and stimulation stopped. This removes the inhibitory influence that, during tetanus, had prevented the relaxation reaction and thus, now, $M_C \rightarrow M$ proceeds. Relaxation in the actual muscle will not begin at the instant the stimulation has stopped, however, since it is reasonable to suppose that the inhibitory reaction products formed by the previous chemical transformations will have accumulated during the tetanus and will not disappear instantaneously. Thus, the relaxation from the plateau of a tetanus will begin some time after cessation of stimulation, and, furthermore, its onset will be gradual, so that the curve for relaxation will be sigmoid. However, the theory indicates that very soon after stimulation stops, the relaxation will be dominated by the simple reaction $M_C \rightarrow M$, and thus the curve for the formation of M will be represented by the function $e^{-k_4 t}$. How closely this is so, in an actual case, is indicated in FIGURE 6, for which $k_4 = 33$. Hence, analyses of such relaxation curves should give the value of k_4. It now appears that the value of k_4 so determined is frequently very near to that of k_1. Thus, the chemical changes characterized by k_1 and k_4 seem to be kinetically similar and this may be helpful in determining the reaction actually involved in the post-contractile relaxation period. At any rate, again choosing a rounded average value, k_4 has been set at 40 sec.$^{-1}$.

It is now necessary to determine the value of a, the tension coefficient of M_R. This may be done by taking advantage of the fact that, after the LR phase has been completed, the rising tension curve must pass through the zero base line in the course of developing the full-blown contraction. Thus, $T = 0$ at a value of t corresponding to L_1 (see FIGURE 1). However, the theory, in its mathematical formula-

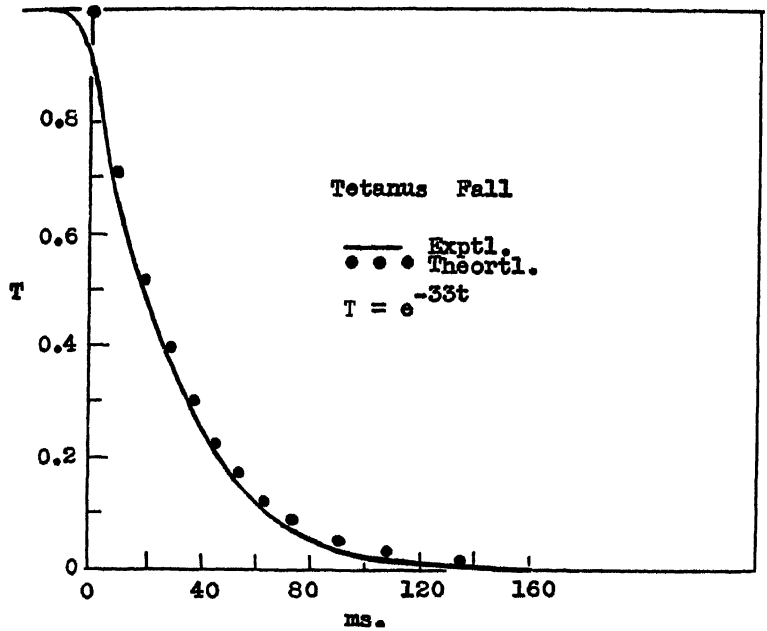


FIGURE 6 Comparison of experimental isometric myogram for the relaxation period of a tetanus and the theoretical values determined from the relation $T = e^{-\alpha t}$. The zero of the time axis is a theoretically extrapolated point. T represents relative tension values

tion, does not take into account the events from the instant of stimulation to the beginning of the LR, and this, analytically, means that the zero of time is shifted from the instant of stimulation to the moment at which the LR is initiated. With this in mind, we note that a typical experimental value of the time at which $T = 0$, is 2.20 ms. Hence, using the general equation for the twitch tension course, we have $T = 0 = f\ (k_1, k_2, k_3, k_4, a, t)$, and all the parameters of this function are known, except a. This equation may be easily solved for a, and its value so determined is 0.0403. Since the values for the constants that enter into the equations for the tension changes are all known, it is now possible to calculate numerically the variation of tension with time under a variety of conditions, and to compare the resultant curves with those obtained under corresponding experimental conditions.

I shall first deal with the twitch, subdividing the discussion into two parts: LR and then the rest of the twitch. In FIGURE 7, note first the equation for the twitch tension course. This covers the complete twitch, *i.e.*, it includes the LR, the contraction, and the post-contractile relaxa-

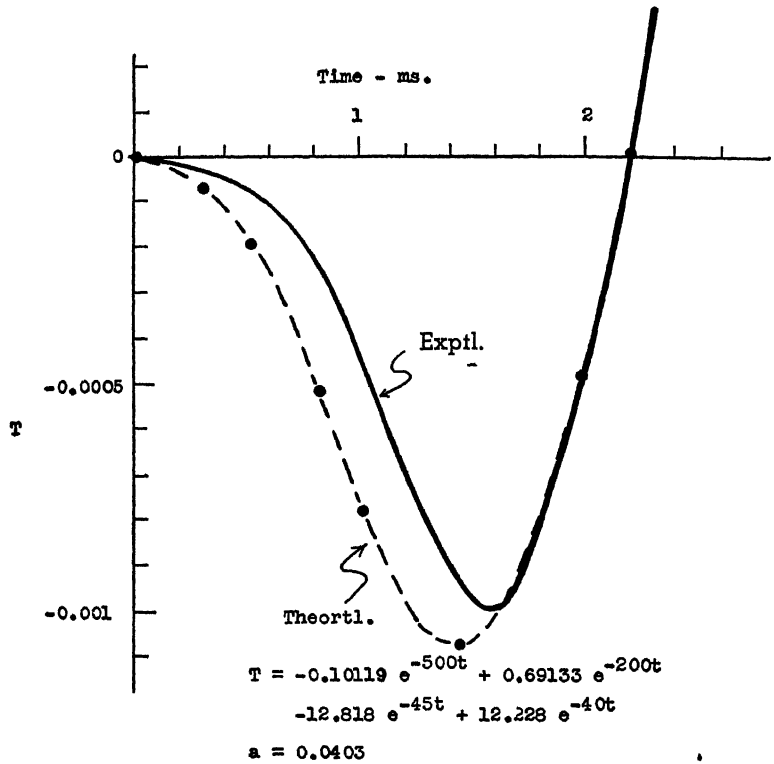


FIGURE 7. Experimental and theoretical LR curves. The theoretical curve is calculated from the expression for T which represents the entire twitch mechanogram determined with $k_1 = 500$ sec.$^{-1}$, $k_2 = 200$ sec.$^{-1}$, $k_3 = 45$ sec.$^{-1}$, $k_4 = 40$ sec.$^{-1}$, and $a = 0.0403$. Maximum tetanus tension = 1. The zero of time is the instant at which the LR begins. The instant of stimulation is at about — 1.2 ms.

tion periods. In the graph are compared the theoretical and the experimental LR curves. The experimental curve is a typical response obtained through stimulation by means of Brown's massive electrodes. The depth of this LR has been set at a value of 1/1000 the peak tetanus tension, this being a rounded experimental average. The dashed line is the theoretical LR curve which has been scaled to the experimental curve only by making the tension equal to zero at 2.2 ms., as has already been indicated in connection with the analytical procedure used for obtaining the value of the constant a.

It will be noted that the theoretical LR curve is a fair approximation to the experimental. Although the down limb of the theoretical curve is ahead of the actual curve, it has essentially the same sigmoid shape, and its depth is only slightly greater than that found in experiment. The rising phases of the two curves are practically identical.

FIGURE 8 is concerned with the main periods of the twitch, the solid line representing the theoretical twitch course. This curve cannot show the LR phase, since the latency events are so short in time and so minute in tension change. However, the twitch curve on the present scale does indicate a definite latent period, as in actual twitches. The general time course of the theoretical twitch is quite like that of the actual isometric myogram of frog skeletal muscle at room temperature for which the various constants have been chosen. This is seen in the following respects: (1) the time of greatest rate of tension rise at about 10 ms.; (2) the time of peak tension at about 35 ms.; and (3) the time to end of the twitch at about 150 ms. Furthermore, the peak tension output is rather correctly given, at least for the frog

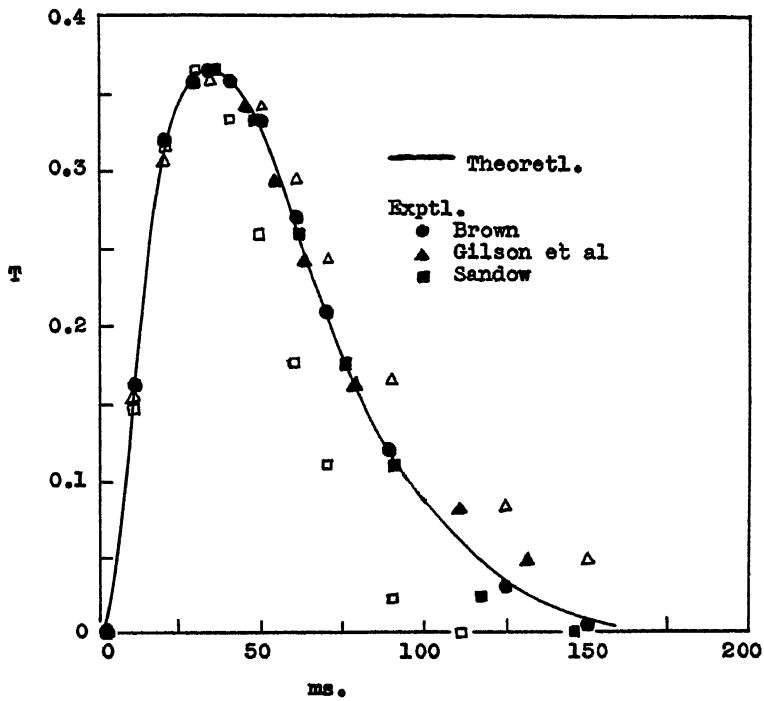


FIGURE 8. Experimental and theoretical twitch curves. Refer to text for further explanation.

sartorius. The average of many experiments in my laboratory shows that the ratio of twitch to tetanus tension is about 0.39. It will be noted that the theoretical value for this ratio is 0.37, in quite good agreement with experiment.

It is now of interest to see how closely the theoretical twitch myogram fits particular experimental curves. The solid circular dots conform to a sartorius twitch at 0.5° C. obtained by Brown. This curve has been scaled to the theoretical curve for height and time at the peak point, and it is evident that, with these adjustments, the agreement between theory and experiment is quite good. The open triangles and squares represent other twitches, as indicated in the legend of the graph, which have been scaled only as to magnitude at peak. Although the theoretical curve agrees with the actual for the contraction period, the divergence is large in the relaxation periods. The calculated curve, however, occupies a median position with respect to these two recorded myograms. Furthermore, these relaxation divergences are much less pronounced when the experimental points, as shown by the solid triangles and squares, are scaled to the theoretical curve by equalizing the times at tension of about 0.1 unit. Thus, generally, and for particular twitches, also in terms of proper scaling, the theoretical twitch is a fairly good approximation to the experimental.

We will now study the tetanus. Relaxation from the full plateau of a tetanus has already been considered in connection with the procedure for evaluating k_4 (see FIGURE 6), and it will be recalled that the experimental curve, except for the brief initial sigmoid part, is quite exactly fitted by the theoretical curve, the equation for which is given simply by $e^{-k_4 t}$. FIGURE 9 illustrates the degree of agreement for the rising phase of the tetanus, the solid line representing the behavior of a sartorius at room temperature in one of our own myograms. The agreement is not as good as desired, but the general course is like that found in experiment. It should be noted, furthermore, that the theoretical equation for the rise of tetanus tension includes an initial LR phase quite like that already demonstrated for twitch. Thus, this curve agrees with the actual LR found associated with the first twitch-like response that initiates the tetanus.

The general agreement that has been demonstrated between the various theoretical predictions and their experimental counterparts suggests that the fundamental assumptions of the theory are, in some measure, essentially correct. This view applies most directly to the following fundamental hypotheses: (1) that, during a twitch response, the contractile material—myosin, or some complex involving this pro-

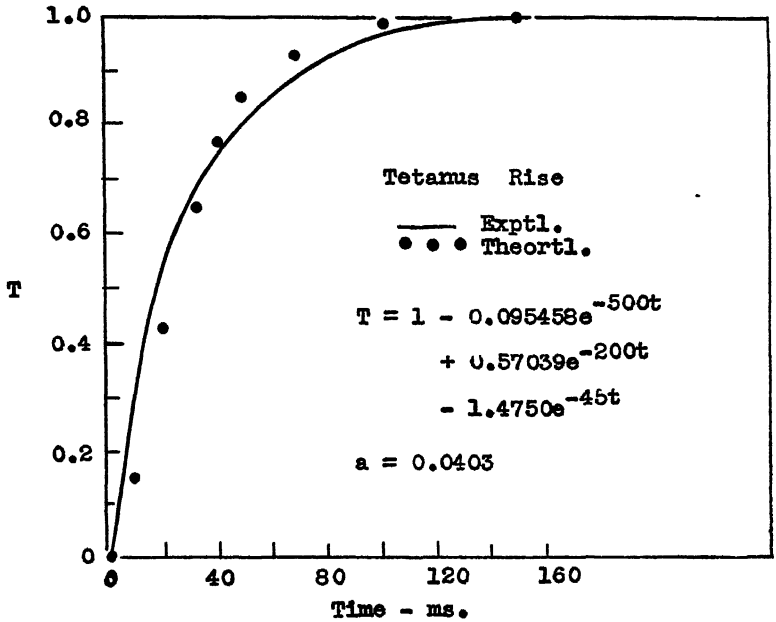


FIGURE 9 Experimental and theoretical curves for rise of tetanus tension.

tein—goes through a series of changes outlined by the reaction chain (I′) which includes four consecutive, approximately first-order reactions; (2) that the tension output as a function of the time is represented by the equation $T = m_0 - am_R$; and (3) that the velocity constants for the basic reactions may be obtained by the methods previously described. Clearly, even if the reaction chain (I′) be accepted as true, the particular chemical reactions that have been assigned to the various transformations of the contractile material may not apply at all. Other chemical changes are quite possible, despite the experimental evidence we have adduced in support of the special role attributed to the interaction between myosin and ATP. For example, it may be that this enzymatic transformation occurs in the form of recovery energization of the myosin, instead of the activity energization our present experimental results suggest. Evidence supporting one or another type of recovery energization is given in several other papers of this conference by Ramsey (1947), Shen (1947), and Steinbach (1947), and specially pertinent in this respect is the speculation of Szent-Györgyi (1946). However, Brown's (1946) contribution to this conference favors activity energization, and his concep-

tion is, in some detail, in agreement with ours. Further discussion of the general point at issue will be found in Sandow (1945a).

Since none of the above references discuss Binkley's theory of mechano-chemical coupling—a form of recovery energization—brief consideration will be given to it here. Binkley (1945) hypothesizes that myosin at rest is a relatively elongated fiber containing neighboring sulfhydryl and phosphorylated hydroxyamino acid side chains. Upon stimulation, contraction results from the condensation of these side chains by the formation of a thio-ether linkage between them and the concurrent splitting-off of phosphate, which releases energy for contraction. Relaxation then occurs in consequence of the reaction of the shortened myosin with ATP which, acting at the thio-ether linkage, splits this bond and, by giving up its first phosphate, restores the original relaxed and energized myosin through reformation of the sulfhydryl and phosphorylated side chains. This scheme has the merit of presenting a definite sequence of chemical reactions, speculative though they be, to account for the mechanical behavior of myosin within the general framework of a recovery energization hypothesis. The theory, however, is fundamentally at fault when certain of its details are examined in terms of present knowledge. Thus, the assumption that myosin at rest is phosphorylated is not in agreement with current evidence. Furthermore, if it nevertheless be assumed that phosphorylated groups do exist in relaxed myosin, it is noteworthy that the particular group that enters into Binkley's mechanism is of the low-energy type. Hence, when this is formed, by interaction of ATP with the thio-ether linkage supposed to exist in contracted myosin, a loss of some 7000 cal. per mol ATP must be accounted for during the relaxation period of a contraction. Although there is heat liberated during relaxation, it does not seem likely that this heat is large enough to account for the large energy loss predicted by Binkley's theory, and, in any case, this heat has been accounted for on other grounds in both isotonic (Hill, 1938) and isometric (Fenn, 1945) contractions. Finally, it is significant that Binkley's scheme depends on the reactivity of — SH groups. Now, iodoacetate is known to inactivate such groups, yet an IAA-poisoned muscle contracts and relaxes, initially, quite normally, in contradiction to the expectation based on Binkley's mechanism. This criticism may be invalidated by the finding of Singer and Barron (1944) that iodoacetamide does not inhibit the enzymatic activity of myosin ATPase. However, in view of the above difficulties, it would seem that Binkley's theory is not an adequate basis for the energization of muscular activity.

Generally speaking, present evidence is thus indecisive regarding the question of activity *versus* recovery energization of myosin. Such success as our theory of mechano-chemical coupling achieves, does not necessarily validate our own assumption of an activity energization; this must be settled, in the final analysis, by relevant experimental data.

Quite apart from the uncertainty characterizing our hypothesis of activity energization, criticism of the theory we propose may be made on other grounds. For example, we have not taken into account the role of muscular viscous forces or the mechanical characteristics of the recording lever system in determining the shape of the mechanical response. These factors, however, may be of only second-order importance. A more serious criticism may be directed against the special manner in which the theory relates the twitch to the tetanus form of the mechanical change. For, despite the existing evidence that each stimulus in a tetanus releases additional contractile material, our mechanism involves a process in which the first stimulus makes potentially available all the material for development of a tetanic response, and the function of the successive intratetanic stimuli is, then, to suppress the relaxation reaction.

It is obvious that only further experimental evidence and, possibly, further extension of the theory can resolve this contradiction. Indeed, much experimentation is still needed to establish the many assumptions that underlie the theory, such as those, for example, concerning the chemical nature, and the time constants, of the reactions which release the activator for myosin-ATPase and account for the combination of these substances. For the future, also, must be left an attempt to coordinate the theory with the work of the Szent-Györgyi group. The present theory may not withstand the impact of new data, and it may thus be relegated to the limbo already rich with previously discarded theories of contraction. In any case, it will have served a useful purpose if it is helpful in suggesting new experimental approaches to the solution of the problem of muscular contraction.

LITERATURE CITED

Bailey, K.
1942. Myosin and adenosine-triphosphatase. Biochem. J. **36**: 121–139.

Binkley, F.
1945. A chemical-mechanical theory of muscular contraction. Science **102**: 477–479.

Brown, D. E. S.
1941. The regulation of energy exchange in contracting muscle. Biol. Symp. **3**: 161–190.

Brown, D. E. S.
1946. The potential role of the myosin-ATPase calcium complex in the activation and energy output of muscle. Unpublished observations, Conference on Muscular Contraction, N. Y. Acad. Sci.

Brown, D. E. S., & F. J. M. Sichel
1936. The isometric contraction of isolated muscle fibers. J. Cell. & Comp. Physiol. **8**: 315–328.

Brown, D., & H. C. Lawler
1946. The activation of myosin-ATPase by pressure in the presence of calcium. Fed. Proc. **II 5**(1): 13.

Buchthal, F., & G. G. Knappeis
1943. Propagation of contraction in the isolated striated muscle fibre. Acta Physiol. Scand. **5**: 256–270.

Dainty, M., A. Kleinzeller, A. S. C. Lawrence, M. Miall, J. Needham, D. M. Needham, & Shih-Chang Shen
1944. Studies on the anomalous viscosity and flow birefringence of protein solutions. III. Changes in these properties of myosin solutions in relation to adenosinetriphosphate and muscular contraction. J. Gen. Physiol. **27**: 355–399.

DuBois, K. P., & V. R. Potter
1943. The assay of animal tissues for respiratory enzymes. III. Adenosinetriphosphatase. J. Biol. Chem. **150**: 185–195.

Engelhardt, W. A.
1941. Enzymatic and mechanical properties of muscle proteins. Adv. Cont. Biol. **14** (ed. 2).
1942. The above article reprinted in translation. Yale J. Biol. & Med. **15**: 21.

Fenn, W. O.
1945. Contractility. Sec. 7 of **R. Höber,** *Physical Chemistry of Cells and Tissues.* Blakiston Co. Philadelphia.

Fischer, E.
1926. Die Zerlegung der Muskelzuckung in Teilfunktionen. III. Die Isometrische Muskelaktion des curarisierten und nicht-curarisierten Sartorius, seine Dehnbarkeit und die Fortpflanzung der Dehnungswelle. Pflüg. Arch. **213**: 352–369.

Gilson, A. S., Jr., G. M. Schoepfle, & S. M. Walker
1947. The time course of tension development in the muscle response. Ann. N. Y. Acad. Sci. **47**(6): 697 ff.

Hill, A. V.
1938. The heat of shortening and the dynamic constants of muscle. Proc. Roy. Soc. London B **126**: 136–195.

Mirsky, A. E.
1937. Contraction of muscle and denaturation of myosin. Proc. Soc. Exp. Biol. & Med. **37**: 157–159.

Ramsey, R. W.
1944. Muscle: Physics. In: *Medical Physics,* ed. Glasser. Chicago.

Ramsey, R. W.
1946. Dynamics of single muscle fibers. Ann. N. Y. Acad. Sci. **47**(6): 675 ff.

Ramsey, R. W., & S. F. Street
1940. The isometric length-tension diagram of isolated skeletal muscle fibers of the frog. J. Cell. & Comp. Physiol. **15**: 11–34.

Rauh, F.
1922. Die Latenzzeit des Muskelelementes. Zeits. Biol. **76**: 25–48.

Ritchie, A. D.
1933. Theories of muscular contraction. J. Physiol. **78**: 322–334.

Sandow, A.
1944. Studies on the latent period of muscular contraction. Method. General properties of latency relaxation. J. Cell. & Comp. Physiol. **24**: 221–256.
1945a. The mechanism of energizing muscular contraction. Trans. N. Y. Acad. Sci. (Ser. II) **7**: 139–151.
1945b. The effect of activity on the latent period of muscular contraction. Ann. N. Y. Acad. Sci. **46**(3): 153–184.
1947. Introduction to the Conference on Muscular Contraction. Ann. N. Y. Acad. Sci. **47**(6): 667 ff.

Schoepfle, G. M., & A. S. Gilson, Jr.
1945. Some observations on the configuration of the free loaded and after loaded muscle twitch. J. Cell. & Comp. Physiol. **26**: 119–130.

Shen, S. C.
1947. The interaction of myosin and adenosinetriphosphate. Ann. N. Y. Acad. Sci. **47**(6): 875 ff.

Sichel, F. J. M.
1941. The relative elasticity of the sarcolemma and of the entire skeletal muscle fiber. Am. J. Physiol. **133**(2): 446–447.

Singer, T. P., & E. S. G. Barron
1944. Effect of sulfhydryl reagents on ATPase activity of myosin. Proc. Soc. Exp. Biol. & Med. **56**: 120–124.

Singher, H. O., & A. Meister
1945. The adenosinetriphosphatase activity of myosin preparations. J. Biol. Chem. **159**: 491–501.

Steinbach, H. B.
1947. Intracellular inorganic ions and muscle action. Ann. N. Y. Acad. Sci. **47**(6): 849 ff.

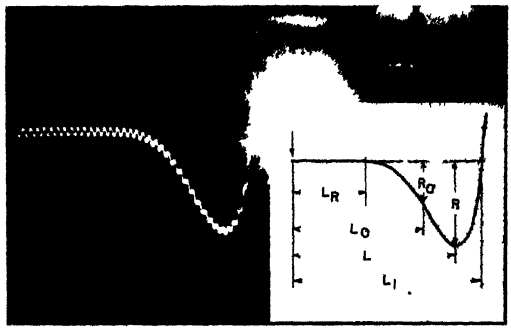
Szent-Györgyi, A.
1946. Contraction and the chemical structure of the muscle fibril. J. Coll. Sci. **1**: 1–19.

PLATE 8

a. Typical oscillographic record of a latent period response of a frog sartorius muscle. The timing modulation is 10,000 cycles/sec. The light bands at the upper right measure the initial and peak developed tension of the contraction. The inset indicates the symbols of the measured latency variables. See text for further explanation.

b. Transversely recorded latency response from Schoepfle & Gilson. See text for explanation.

c. Transversely recorded latency response from a frog sartorius. Pickup stylus is the cathode, 10,000 cycles/sec. timing wave. Light band at right indicates initial and contraction transverse tension.



a

B

A

C

b

c

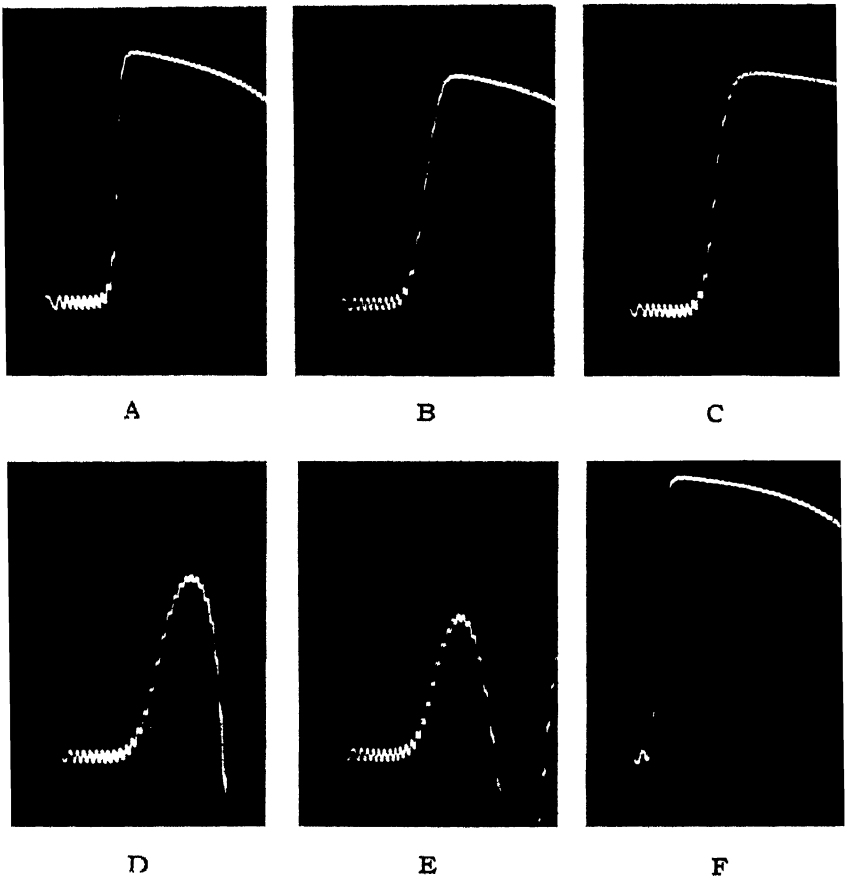

A
B
C
D
E
F

PLATE 9

Representative oscillographic records of displacement of the upper end of a frog sartorius muscle in response to a suddenly applied downward pull applied to the lower end. 10,000 cycles/sec. timing wave. Records A through E correspond to progressively slower rates of pull. Record F registers the effect of the fastest rate of pull when applied directly to the stylus of the pickup.